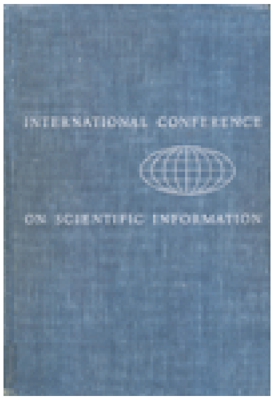


Proceedings of the International Conference on Scientific Information -- Two Volumes

Sponsors of the Conference: National Science Foundation, National Academy of Sciences, American Documentation Institute, National Research Council

ISBN: 0-309-57421-8, 1662 pages, 6x9, (1959)

Proceedings of the International Conference on Scientific Information

WASHINGTON, D.C. · NOVEMBER 16–21 · 1958
IN TWO VOLUMES

Sponsors of the Conference: National Science Foundation
National Academy of Sciences—National Research Council
American Documentation Institute

National Academy of Sciences—National Research Council
Washington, D.C. · 1959

Library of Congress Catalog Card Number 59–60045

Printed in the United States of America

ALBERTO F.THOMPSON

DECEMBER 1, 1907 · JUNE 18, 1957

IN MEMORIAM

THE CONFERENCE owes more perhaps to Dr. Alberto F.Thompson than to any other individual, for he transformed the initial conception into a plan that others finally carried out. As Head of the Office of Scientific Information of the National Science Foundation he was deeply involved in the planning of the Conference, possibly too deeply, for he gave himself with boundless enthusiasm to all that interested him, regardless of limitations of time and health.

An organic chemist, Dr. Thompson received his Ph.D. degree at Harvard, did post-graduate work at the University of Munich, and taught at the University of Minnesota and The Massachusetts Institute of Technology. As a major in the Manhattan District of the US Corps of Engineers during World War II, he worked on the development of the atomic bomb. He became Chief of the Technical Information Service of the US Atomic Energy Commission and in November, 1955, he joined the staff of the National Science Foundation. Among his achievements were publication of the *National Nuclear Energy Series* and the establishment of *Nuclear Science Abstracts*.

His infectious good humor and the brilliant range of his interests (from limericks and model railroads to the works of Mozart and the cultivation of roses) won the affection of all, even of those who disagreed with him.

Too energetic and too wise to see science in terms less than international, he saw the information problem on the same scale; yet he searched always for the most effective immediate measures. Operations research on the flow of scientific information received strong encouragement from him: he was deeply interested in mechanical translation and electronic data processing systems. At the same time, he had utmost respect for the physically simple retrieval systems.

Alberto Thompson's expectations for the Conference combined high hopes with New England practicality. We hope that the Conference succeeded in achieving what he would have wished: to inspire us with the vision of the future without letting us forget the realities of the present.

GILBERT W.KING

VOLUME ONE

VOLUME TWO

CONFERENCE ORGANIZATION

Conference Committee

Wallace W.Atwood, Jr., Chairman and NAS-NRC Representative
Burton W.Adkinson, NSF Representative
Milton O.Lee, ADI Representative
Charles I.Campbell, Program Committee
Henry J.Dubester, Local Arrangements
John C.Green, Exhibits
Mary McC.Sheppard, Secretary

Program Committee

Charles I.Campbell, Chairman
Helen L.Brownson, Area 1
Dwight E.Gray, Area 2
Joseph Hilsenrath, Area 3
Mary Elizabeth Stevens, Area 4
H.P.Luhn, Area 5
Lawrence F.Buckland, Area 6
Frank B.Rogers, Area 7

Discussion Panel Chairmen

Area 1 Philip M.Morse, Department of Physics, Massachusetts Institute of Technology, Cambridge, Mass.

Area 2 Elmer Hutchisson, American Institute of Physics, New York, N.Y.

Area 3 Alexander King, European Productivity Agency, Paris, France

Area 4 Eric de Grolier, Centre Français D'Échanges et de Documentation Techniques, Milan, Italy

Area 5 Gilbert W.King, IBM Research Center, Yorktown Heights, N.Y.

Area 6 John W.Tukey, Department of Mathematics, Princeton University, Princeton, N.J.

Area 7 Verner W.Clapp, Council on Library Resources, Inc., Washington, D.C.

Local Arrangements Committee Henry J.Dubester, Chairman

Marion E.Bonniwell · Saul Herner · Rita G.Liepina

Wyvona A.Lane · Madeline M.Berry

Exhibits Committee John C.Green, Chairman

Eugene E.Miller · Gerald J.Sophar · Don D.Andrews · Isaac Fleischmann

PREFACE

ON BEHALF OF ALL THOSE who for the past three years have devoted considerable time and effort in preparation for the International Conference on Scientific Information, it is my privilege to present herewith the Proceedings. Certain members of the American Documentation Institute, among them Milton O.Lee, originally conceived the idea for this type of conference. They wanted to bring together on an international level scientists and information specialists for discussion of current research progress and problems concerned primarily with the storage and retrieval of scientific information. Ultimately these aims and ideas were developed until there resulted this Conference, jointly sponsored by the American Documentation Institute, the National Science Foundation, and the National Academy of Sciences—National Research Council.

The American Documentation Institute is a private organization, supported by membership dues. After a modest beginning in 1937 it grew in size and stature until today its membership includes some 300 individuals professionally engaged in working with information and documentation in one capacity or another. In 1947 the Institute became the United States national member of the International Federation for Documentation.

The National Science Foundation, an independent agency of the Federal Government, was established in 1950 by Act of Congress. Its main functions are to support basic research and education in the sciences and to foster the exchange of scientific information. The chief executive officer of the Foundation is the Director. Final responsibility for establishing Foundation policy lies with the 24-member National Science Board whose distinguished members are appointed by the President of the United States with the approval of the Senate. The Foundation is playing an increasingly significant role in strengthening the scientific capabilities of the country.

The National Academy of Sciences is a private, non-profit organization established in 1863 and dedicated to the furtherance of science for the general welfare. Its membership is comprised of more than 550 leading scientists of this country. Its congressional charter provides that the Academy advise the Government on scientific matters. In 1918 the National Research Council was established by executive order of the President of the United States as part of the National Academy of Sciences, and has since given the organization its present character and popular name of the Academy—Research Council.

Appropriately, the opening session of the Conference on 16 November included welcoming remarks by the representatives of the three sponsoring bodies. Milton O.Lee greeted the delegates on behalf of the American Documentation Institute and reviewed the way the Conference plan had originated, developed, and matured. Alan T.Waterman, the Director, spoke on behalf of the National Science Foundation. The final welcome was extended by Detlev W.Bronk, President of the National Academy of Sciences. In concluding his remarks Dr. Bronk expressed gratitude to the Royal Society for having had the vision to sponsor a significant conference on scientific information in 1948. Thus the stage was set for the address by Sir Lindor Brown, which is reproduced elsewhere in these volumes.

The Conference Banquet on 19 November marked the midway point for the Conference. On this occasion Alexander King, President of the International Federation for Documentation, was toastmaster and introduced the speakers.

Dr. Waterman, who spoke first, said that during the ten years which had passed since the Royal Society Conference, an increasing recognition had developed of the problems and the importance of scientific information. He reviewed briefly the efforts of the United States Government to meet scientific information needs and, more specifically, what the National Science Foundation was prepared to do in this important area.

The second speaker was Herman Henkle, President of the American Documentation Institute. Dr. Henkle's informal remarks stressed the immediacy of the roots Americans have in overseas lands and in the cultures of those lands.

The principal address was given by Dr. Bronk who drew attention to the need for a synthesis of knowledge at a time when specialization grows ever more common and more narrow and so tends to erect new barriers to understanding. The solution to this problem, Dr. Bronk said, must lie in developing a broad awareness while cultivating one's own special knowledge. He believed that this, plus individual integrity, was the best assurance against the danger that scientists might find themselves cut off from one another by each one's exclusive concentration on his own field of interest.

Having given this very brief summary of the opening of the Conference and the banquet program, I commend to you the Conference reports prepared by the panel chairmen and which constitute a major and significant portion of these Proceedings.

Although it would be impossible to name all those whose special skills were devoted at one time or another to insuring the success of the Conference, I would like to take this occasion to acknowledge their many contributions.

And especially I express appreciation to my colleagues on the Conference Committee who carried the responsibilities of Conference planning and organization: to Charles I.Campbell, chairman of the Program Committee; to Henry Dubester, chairman of the Local Arrangements Committee; to John Green, chairman of the Exhibits Committee; and to our talented Executive Secretary, Mrs. Mary McC.Sheppard, whose enthusiasm and boundless energy gave courage and support to all of us throughout the 3 years of Conference activity. To these persons and to the large corps of volunteers who served as members of committees, I extend the sincere thanks of the sponsoring organizations.

It is our hope that this Conference will stimulate further research and closer cooperation among those who are attempting to cope with the problems involved in making scientific information easily and rapidly available. We also hope that these Proceedings will be of value to the many hundreds who were unable to attend the discussion sessions of the Conference.

WALLACE W.ATWOOD, JR.

INTRODUCTION

THE PROCEEDINGS of the International Conference on Scientific Information, which are published here, will be better understood if it is explained how the Conference aim was defined and how the program was arranged to advance that aim.

During the spring and summer of 1956, an informal Preliminary Planning Committee met nearly every week to define the scope of the Conference and to devise a plan for carrying it out. Chairman of the Preliminary Planning Committee was Milton O.Lee, American Physiological Society. Members included: Scott Adams, National Institutes of Health; Samuel Alexander, National Bureau of Standards; Robert F.Bray, The Library of Congress; Helen L.Brownson, National Science Foundation; Charles I.Campbell, National Academy of Sciences-National Research Council; Verner W.Clapp, Council on Library Resources, Inc.; J.E.Cummins, Australian Scientific Liaison Office; Dwight E.Gray, National Science Foundation; John C.Green, Department of Commerce; Joseph Hilsenrath, National Bureau of Standards; William T.Mason, Department of Commerce; Frank B.Rogers, National Library of Medicine; Mary Elizabeth Stevens, National Bureau of Standards; and Mortimer Taube, Documentation, Inc. A provisional Secretariat was established at this time with Alberto F.Thompson of the National Science Foundation as Executive Secretary and Mary McC.Sheppard of the Academy-Research Council as his assistant. Many others from this country and abroad met with the Committee at various times to give counsel and guidance.

After preliminary plans and working documents were developed, an *ad hoc* committee composed of 50 distinguished scientists and information specialists, under the chairmanship of Warren Weaver of the Rockefeller Foundation, was convened on November 11, 1956, at the request of the Academy-Research Council, to review the proposed Conference plans, its aims and scope, and to determine whether such a Conference was warranted. At the recommendation of this *ad hoc* committee, planning for the Conference proceeded. The proposed content of each of the seven areas of discussion in the Conference was outlined in detail, and the following criteria for acceptable papers were established:

1 Papers will deal with work that has not been published or presented at any open meeting. Work will be considered to have been published if it has been reproduced

for general distribution in any form or if copies have been deposited in libraries where they are available to the public.

2 Papers will be directed to specialists in the field covered. Only sufficient background information will be included to serve as an adequate framework for new work described in the papers. More general background material will be indicated by references.

3 Papers dealing with systems and methods will describe these at length only when they have not been described previously. If new methods or systems are involved, these will be described in sufficient detail to enable other qualified workers to duplicate the procedures and the results. There will be sufficient information to enable qualified readers to judge the validity of results in objective terms.

4 Theoretical papers will clearly explain the factual basis from which theoretical conclusions have been drawn and will point the way to experimental methods of verifying predictions which follow from such theoretical conclusions.

These criteria, together with the definitions of the Discussion Areas, which will be found at the opening of each section in these volumes, were provided to all prospective authors.

Early in 1957, a formal policy committee was created with Milton O.Lee, representing the American Documentation Institute; Wallace W.Atwood, Jr., representing the Academy-Research Council; Alberto F.Thompson, representing the National Science Foundation; and Eugene Power and J.E.Cummins named as members-at-large. Also established at this time was a Program Committee with responsibility for reviewing and selecting papers in accordance with the scope and criteria for papers, for appointing discussion panel members, and for making arrangements for the Conference program proper. Charles I.Campbell, The Rockefeller Institute, was named Chairman. Other members of the Committee, selected from the preliminary planning group, and their respective areas of responsibility were: Area 1, Helen L.Brownson; Area 2, Dwight E.Gray; Area 3, Joseph Hilsenrath; Area 4, Mary Elizabeth Stevens; and Area 7, Frank B.Rogers. Two new members were added: H.P. Luhn of the IBM Research Center who accepted the responsibility for Area 5, and Lawrence F.Buckland of Itek Corp., for Area 6. Miss Madeline M.Berry of the National Science Foundation was of very great assistance to the committee, especially in developing the program for Area 5.

All those connected with the Conference were saddened by the death of Alberto F.Thompson in June, 1957. During the reorganization which followed, a Conference Committee was established early in 1958 with Wallace W.Atwood, Jr., of the Academy-Research Council becoming Chairman and Mary McC.Sheppard continuing as Secretary. Thomas O.Jones, Acting Head of the Office of Scientific Information, provided valuable help as a Committee member from June to December, 1957, when Burton W.Adkinson became Head of the Office of Scientific Information and thus became the Foundation's representative on the ICSI policy committee. About this time Eugene Power resigned because of the pressure of other activities, and J.E.Cummins resigned when he accepted a position with the International Atomic Energy Agency in Vienna. Others named to the Committee in 1958 were: John C. Green, Department of Commerce, in charge of exhibits; and Henry J.Dubester, The Library of Congress, in charge of local arrangements. The Committee structure thereafter remained unchanged.

Because of the narrowly defined scope of the Conference and the intention rigorously to select contributions, it was believed to be wise to consider outlines of proposed papers well in advance of the preparation of papers themselves. We hoped in this way to avoid at least part of the grief of declining to accept papers that had been, in a sense, solicited. During 1957, therefore, an immense amount of correspondence was carried on by the members of the Program Committee and the Secretariat with somewhat under a thousand potential authors of papers in nearly every country of the world. All decisions on papers were taken by the Committee jointly, though we often sought the guidance of referees. We were forced in some cases to decline very sound contributions that concentrated on aspects of the scientific problem that had been excluded from the program explicitly or implicitly. From the approximately 150 papers that were given formal consideration, 75 papers were selected.

These papers served as stimulating and valuable points of departure for the discussions of the Conference. We may hope that through their publication here they may provide a basis for further progress in research throughout the world.

CHARLES I.CAMPBELL

CONTENTS

VOLUME TWO

AREA 5 *Organization of Information for Storage and Retrospective Search: Intellectual problems and equipment considerations in the design of new systems*

OPENING SESSION ADDRESS

OPENING SESSION ADDRESS

SIR LINDOR BROWN

The invitation to give an opening address at this International Conference on Scientific Information reached me at a time when I happened to be reading a description of the greatest bore in English literature, a description that appears appropriately in a book written jointly by a Scotsman and an American, Robert Louis Stevenson and Lloyd Osborne. It is said of Uncle Joseph Finsbury that “A taste for general information, not promptly checked, had soon begun to sap his manhood. There is no passion more debilitating to the mind, unless perhaps it be that itch of public speaking which it not infrequently accompanies or begets. The two were conjoined in the case of Joseph; the acute stage of this double malady, that in which the patient delivers gratuitous lectures, soon declared itself with severity, and not many years had passed over his head before he would have travelled 30 miles to address an infant school.” Tonight I have travelled more than 30 miles, and visual observation suggests that I am not addressing an infant school. If you will but change “general information” into “scientific information” you will see what a shocking coincidence this was and how I had to search my mind to find justification for my presence before you tonight. The simplest explanation is that I am at the moment the Senior Secretary of the Royal Society of London, and the Royal Society held almost exactly 10 years ago the first Conference of international status devoted solely to the subject of Scientific Information. The Conference was not in the strict sense international since it arose directly from the deliberations two years previously of a meeting of Scientists of the British Commonwealth, what we then had the temerity to call the British Empire Scientific Conference. The Royal Society Scientific Information Conference therefore was attended by representatives from all the Dominions, and, through a very fortunate chance, or perhaps as a result of a bland disregard of the Declaration of Independence or, and this is the more likely course, through the wisdom and foresight of my predecessors, one of whom, Sir Alfred Egerton, is in this room tonight, it included a representative from the Library of Congress, it included Dr. Murray Luck and it included no less a person than one Dr. Detlev W.Bronk, then

SIR LINDOR BROWN, C.B.E. Senior Secretary of the Royal Society of London.

described as “Foreign Secretary of the National Academy of Sciences, Washington.” This goes a long way towards explaining not only my presence before you, but the very existence of this Conference.

We are fortunate on this present occasion of knowing in detail what the Conference is about before it meets, and I should like to take this opportunity of congratulating the Conference Committee on the boldness of conception and the enterprise that they have shown in planning their proceedings in this way. This seems to me to be a daring experiment, and they are putting much faith in the skill and efficiency of the Panel Chairmen. But they have been well chosen and I am sure that they have already taken tremendous pains in organizing their discussions.

Now let us compare the advance literature of the present Conference with the report of its predecessor of ten years ago. One difference is obvious and inescapable, the final, full report of the 1948 Conference weighs 1 Kg. The weight of the advance information *only* for the present one is 2.72 Kg. If these documents represented the publications of the applicants for a post, there is no doubt at all but that the second, the younger applicant would be successful. Now, as we have tacitly assumed the functions of a selection committee, let us take the unusual step of looking inside these two anthologies; not taking the extreme, and perhaps invidious step of assessing the relative values of individual contributions, but rather comparing the subject matter of their contents. The Royal Society's Conference had 17 “Working Parties”—it was the era of the “Working Party”—but I will not weary you with a catalogue of their terms of reference. One group of five working parties dealt with the problem of the raw material of Scientific Information—Scientific publication format, editorial policy, subject grouping, general organization and delays in publication. Then there was a group dealing with Abstracting and another with Reviews and Annual Reports. Very considerable activity was shown by the groups of working parties dealing with classification and indexing and the training of information officers.

Our conference in 1948 covered then the whole subject of scientific information from begetting to burial, or as we say in the Welfare State, from the cradle to the grave. It has been the intention of the organizers of this present Conference to limit its interests to the later phases of the life cycle of information— to storage and retrieval. I dare not say to burial and disinterring. It is a measure of the growth of interest in this subject over the last 10 years that this relatively small fraction of the whole scientific information problem should have produced a conference eagerly attended by nearly 1000 delegates with, if my information is correct, an equal number clamouring unavailingly for admission. The importance of the subject to Government and to industry is clearly indicated by the formidable list of sponsors that heads the volume of pre-circulated papers.

I, as you well know, am no expert in this subject, I am here to that extent under false pretences. To me the entirely staggering feature of this Conference is that quietly, surreptitiously there has grown up a new science with its own language, its own techniques and its own leaders. Those leaders, from all parts of the world are gathered here tonight, and I still feel that it is something of an impertinence for me to speak to you. Part of the reason for this feeling is that your development has been so great and has proceeded at such a pace that the ordinary scientist, shut away in his laboratory, may not understand your language and so may fail to appreciate your aims and the benefits to himself that your work may bring. This indeed is a failure of communication by those whose primary function is to communicate. I work at the begetting end of the life cycle of scientific information, you work at a distant point in that cycle where you preserve the encysted spores of knowledge and revitalize them on demand. But I maintain that it is a cycle and unless there is complete and effective collaboration at all stages of the cycle, the usefulness of your work, no less than mine, will be diminished.

It is no part of my duty tonight to attempt to teach you your business. Indeed, my very ignorance precludes such an attempt but I want to try to look at the problem of scientific information as a whole. I feel, as everyone connected with science must feel, a little sad that specialization has proceeded so far that we can no longer appreciate the beauty and significance of the work of others in fields only a little removed from our own. I use the word beauty advisedly, because there is an aesthetic quality about good scientific work that contributes greatly to our understanding of it and to the inspiration that it gives us. May I make a small plea to those who supply and to those who use scientific information not to refine their means of supply and their system of use to the extent that they lose this aesthetic quality. If it is lost science will suffer, and we must not lose sight of our objective, our ultimate goal, the improvement of our knowledge of nature and natural laws. Our striving to this goal has now led to the necessity of the development of another speciality, the science of information, and I am a little uneasy lest this specialization may lead the makers of information too far apart from you, the suppliers. There seem to my simple mind to be at the most only three parties in the cycle of information—the maker, the storer and supplier, and the user who is often the maker also; unless these parties work together we shall have failed in our objective, the advancement of knowledge.

I cannot, I am afraid, look at these problems from your point of view because I am ignorant of your specialty, but I confess to a certain uneasiness from the point of view of the user, or of the maker of information. The rate of increase in production of scientific papers is so high that one must face the prospect of a swamping and ultimate failure of information storage and retrieval; for one reason if for no other, a good information service provides positive feed-back and automatically increases the production of papers. Is there anything that we can do about this? Yes, I think there is, but it can be successful only if it is a joint effort by the makers of information and those who store and retrieve.

There can be no doubt in the mind of anyone dealing with scientific literature that much need never be produced and of that which justifies its appearance, 80 per cent could be improved by drastic reduction in length and by clear writing. We all know the economic pressures that give rise to this overproduction of sloppy work. We all know that appointments and promotions depend on quantity rather than quality of work and that much the same criteria are used in assessing the success or otherwise of a research institute or organization. Everyone who has made a discovery has succumbed to the temptation of making a preliminary announcement in *Science* or *Nature* or *Naturwissenschaften,* or in the Proceedings of a specialist society. And now many research organizations are making a practice of distributing widely, advance abstracts and progress reports which anticipate and extensively duplicate the final, comprehensive publication. This regrettable practice is perhaps forgivable in a subject of burning and immediate importance and takes the place of the polite correspondence between savants which was such an attractive feature of the centuries preceding this; but is it forgivable when it spreads more widely, when it is used for priority claims and becomes an encumbrance upon the information services? I feel quite seriously that the time has come for the makers of information to use restraint. Now what do I mean by restraint? I mean a little simple self denial by scientists; let them publish the results of their work once and once only, let us see a well written account in one Journal instead of *exempli gratia* a cyclostyled laboratory preview, a letter to *Nature,* a paper, a lecture given to the Literary and Philosophical Society of Little Puddlebury and printed in its “Proceedings”; the verbatim account of a Symposium at Sulphur Creek, Colorado, and perhaps also a review article ostensibly of the subject but in fact of the author's own work. The Rake's Progress is not yet over because as the author is recovering from his travels by doing a little work in his laboratory and is happily on the track of something new and exciting, he is bedevilled by a request for an article to appear in the *Festschrift* to celebrate the 75th anniversary of the birth of Professor Weissnichtwer, and to save time and trouble produces a rehash of his original paper, clothing the bare discovery in a modest figleaf and twining a vine or two in its hair. All this is very fine for the “Ego” of the author, it may not be bad for his bank balance, but it clogs the information machine and bewilders the expert with the changing facets of a simple discovery presented in a multiplicity of guises.

All I ask is for a little self restraint by authors, but I have little hope of it being fulfilled without strong pressure from outside. Would it not help a little if societies organizing meetings decided to abandon the publication of Proceedings, if those organizing lectures allowed the speaker the privilege of *not* publishing his works and if the organizers of symposia (not, of course, of conferences) refrained from churning out the turgid lucubrations of their participants.

I have suggested that the makers of information might contribute to the easing of your task by a voluntary discipline over publication. But that is a question of *quantity* of information. What about *quality*? Here I know that I am on dangerous ground. As I understand it, the expert on information must, by the very nature of his task, treat all scientific papers as of equal value. The user, as a result of experience, prejudice, or ignorance, treats as significant only a small amount of the information to which he has access, and he still believes that there are good papers and bad papers. You, on the other hand, may well say that a bad paper may contain uncut gems of information or the purest gold of knowledge alloyed and hidden in its depths, but have you really the time, the energy, the money to spend acting as cutters, as polishers, and as refiners of the uncouth products of the incompetent writer? I think not. However we improve the raw material of information there will still remain poor papers, unnecessary papers, trivial papers, and repetitious papers. What are we to do with them? What is to be stored? Is it the fruit of the tree of knowledge or is it the fallen leaves? For my part I should be happy if the fruit only were preserved; as I cannot conceive of any collecting and storage system adequate to cope with the world's output of scientific information as it grows at present, and I view with only a little compunction the prospect of the loss of minor contributions to knowledge provided that the ripe fruit can be preserved. The few fallen leaves of my own work are already losing their individuality as they dissolve into the rich compost of the background knowledge of the rising generation. Their individuality is lost, but they have contributed a little to the vigour of the pushing young shoots at the top of the tree—and of course I have a fruit or two up my sleeve still! But here we have the difficulty, who is to decide what is the fruit and what is the fallen leaf. Can any information service work on an eclectic principle? I think that it can, but only with the full cooperation of the user and the maker of information. If we can persuade the maker to make fewer, but better bricks and if we can persuade him to collaborate with you in casting out the flawed and crumbling stock, we shall build a pyramid of knowledge which will stand and will not fall about our ears with the curse of Babel.

So much for my hobby horse—the unity of scientific endeavour. You are going to have the opportunity of riding yours to your hearts' content in the course of the following week, so you must allow me a passade or two on my own. But my hobby horse is tired, so is its rider and so are you and I must stop, but before I do so I must say a word or two on behalf of us all to the organizers of this Conference. First and foremost to Dr. Atwood and the Conference Committee who seized on a bold idea and developed it swiftly and logically, secondly to the Programme Committee whose labours have produced the remarkable volume that I have already referred to. Finally, may I on behalf of all your guests from overseas say how delighted we are to be here and how great is our determination to give this Conference the success that it deserves.

AREA 1

LITERATURE AND REFERENCE NEEDS OF SCIENTISTS

Knowledge now available and
methods of ascertaining requirements

AREA ORGANIZATION

Authors of Papers

Members of Discussion Panel

Chairman: PHILIP M.MORSE, Department of Physics, Massachusetts Institute of Technology, Cambridge, Mass.

LEONARD CARMICHAEL, Smithsonian Institution, Washington, D.C.

SIR ALFRED EGERTON, Salters Institute of Industrial Chemistry, London, England

HERMAN H.FUSSLER, The University of Chicago Library, Chicago, Ill.

CHAUNCEY D.LEAKE, Department of Pharmacology, The Ohio State University, Columbus, Ohio

WILLARD F.LIBBY, United States Atomic Energy Commission, Washington, D.C.

RALPH R.SHAW, Graduate School of Library Service, Rutgers University, New Brunswick, N.J.

PROPOSED SCOPE OF AREA 1

IN ORDER TO improve the dissemination of scientific information and to design more effective reference tools and services, we need to have a more complete understanding of the weaknesses and strengths of the present pattern of scientific communication and, in particular, of the unfilled or inadequately filled needs of scientists for information. It would be helpful first to have a summary of the views of a number of representative scientists in different fields concerning their own information requirements and problems; and then to review the information yielded by studies of the use of scientific literature and by surveys of various types and to attempt to identify some of the needs suggested by the results of such studies. It is to be expected, however, that a review of past and present studies will point up the need for much additional and more objective knowledge about information practices and needs. Methods by which this knowledge can best be obtained should be considered, both by evaluating the results obtained with the methods that have been tried and by discussing the potential usefulness of other methods that have yet to be applied to this problem.[1]

A few examples of the studies that have been made will serve to illustrate the methods that have been used to date. Dwight Gray and Bentley Glass have made questionnaire surveys in the fields of physics and biology respectively in order to determine how abstracting services are used, the degree of satisfaction with present services, and the express desires of scientists for improved services. As a part of the recent survey of the physiological sciences, the Survey Research Center of the University of Michigan included in a widely distributed questionnaire questions concerning the use of journals and bibliographical publications and the extent to which physiologists believe they are able to keep up with developments in their fields of specialization.

Interviews have been used by Saul Herner in an effort to discover from research scientists just how they look for and obtain information, what publications and library services they use, and the extent to which they do their own searching or rely on the help of librarians and information specialists.

A number of studies have made use of the “reference counting” method. Perhaps the best known of these was a study by Herman H.Fussler to determine

[1] The proposed scope of the Conference Area, as shown here, was prepared during the Spring and Summer of 1956 and provided to all potential contributors as a guide to the aims of the Conference.

the characteristics of the research literature used by chemists and physicists in the United States, the forms and national origins of the literature used, the importance of various subject fields, the temporal span, and the most important journals for each field. Most studies of this type purport to identify the important publications. Estelle Brodman, on the other hand, has attempted to assess the validity of the "reference counting" method by asking a number of individual investigators to rate the publications most useful to them in order of their importance. She found little or no correlation between the results obtained by means of these two approaches.

A "current use" study was made by D.J.Urquhart, prior to the Royal Society Scientific Information Conference, in which the users of periodicals and books circulated by a science library were asked to fill out a brief questionnaire explaining where they had learned of the publication, whether it contained the desired information, and for what the information was needed.

The diary method was first used by J.D.Bernal in a study reported to the Royal Society Scientific Information Conference. He asked working scientists to keep detailed records of what they read, why they read it, and what use they made of the information. The validity of the method has recently been investigated in a pilot study at the Forest Products Laboratory, Madison, Wisconsin, under the direction of Ralph R.Shaw.

D.B.Hertz and A.H.Rubenstein have investigated communication in team research, including the flow of information among members of research groups and the use of outside sources of information. Communication difficulties experienced by the groups were studied by means of interviews. The report of this investigation includes a discussion of methodology, the value of the case-study approach versus short data-gathering visits to organizations being studied.

As for current projects that are trying still different methods, Russell Ackoff, Case Institute of Technology, has directed a study of the feasibility of using the techniques of operations research to obtain objective data concerning the role of information in the research process and the way in which scientists locate and use information. A full-scale study is now being planned that will employ a technique of observation to determine how much time is spent with the literature and which primary and secondary publications are used most extensively. Charles Y.Glock, Bureau of Applied Social Research, Columbia University, is initiating a project to test the value of interviews in depth conducted by skilled interviewers in eliciting from scientists on the staff of Columbia University information concerning their satisfactions and dissatisfactions with the present situation and any problems they have in keeping up with recent advances and in locating specific information when needed.

A summary or, if possible, a synthesis of the results of all these and other studies would bring out the areas of agreement and of disagreement among them. For example, Urquhart, in his study of the current use of library materials, found that abstracts were the main source of information about the publications requested. Shaw, on the other hand, in his recent study based on diary records, found that abstracts stimulated a surprisingly small percentage of the requests for publications. The two studies were carried out in quite different environments, and undoubtedly there was some difference in the subject fields in which the two groups of scientists were working. Nevertheless, this difference in the findings suggests an area for further exploration.

Perhaps the greatest need is for methods that will produce truly objective data that will help to define the current information problems and place them in proper perspective. Important as it is to have the views of research scientists about the adequacy of our present services, they themselves are usually not in a position to know how good a job an information service is doing unless they have had the time to study it carefully. For example, Glass found that 87 per cent of the biologists who participated in his survey regarded the coverage of *Biological Abstracts* as "satisfactory." Glass comments, "This opinion can only arise from a gross misconception of the true state of affairs…. It is clear that users…have little idea of the extent of the gaps and omissions in the abstracting of the periodicals supposed to be covered, or of the vast majority of scientific periodicals with biological material that are not covered at all."

Numerous suggestions have been made for other types of studies that might add to our understanding of information problems and point the way toward improvements most urgently needed, such as:

1 Case studies of actual research projects to determine the role of scientific literature in research, the time given to it, the way it is used, etc.
2 Collection of data on specific instances in which time and money were wasted because information was not readily available, in particular, instances of undesirable duplication of research.
3 Study of reference questions put to information centers to determine just what information the scientists are seeking in specific instances; also further study of "unanswered questions." A British study of questions to which researchers could not find answers suggested that the subject indexes to our abstract journals are inadequate.
4 Studies of the results obtained and changes in the work habits of scientists in situations where skilled information assistance and mechanized searching services are provided.
5 Investigation of the efficiency with which individual scientists use the available facilities to keep up with advances in their own and related fields and

to locate specific data when they need it. A corollary study might inquire how the best informed persons in their respective fields manage to keep up with all that is reported in the literature and to keep track of what has been done in previous years.

6 A psychological study to determine just how persons get information from reading.

7 Exploration of the validity of the hypothesis that the creativity of a scientist is related to the breadth as well as the depth of his knowledge.

8 Exploration of the amount of time it is worthwhile to spend in a literature search in various special fields.

9 Assignment of the same research problem to two different groups, one to have skilled assistance with the literature and the other no special assistance, in order to compare the courses of the work of the two groups.

It is probably correct to say that up until now scientists have been concerned for the most part with maintaining and improving their journals of primary publication and with obtaining better abstracts, indexes, and other bibliographic services, with considerable discussion in recent years of the desirability of faster, mechanized search procedures. It may be, however, that if we had a more complete understanding of the information requirements of research scientists, we would place more emphasis on another aspect of the total picture, such as the preparation of more first-class reviews that synthesize new knowledge and draw attention to the areas where more work is needed. As the volume of the scientific literature continues to increase and as bibliographic and search processes are made more efficient and comprehensive, the quantity of relevant literature on any one topic will become larger. The sheer time required even to scan large numbers of publications identified in a literature search suggests that it is becoming more and more important to find means of reducing to manageable proportions the large quantities of information scattered in many different sources. The problem of reviews will undoubtedly be discussed in Areas 2 and 3, but in this portion of the agenda it might be fruitful to consider requirements for review publications or for other means of "data reduction;" and then in the succeeding areas to discuss ways and means of encouraging their preparation and possible experiments to test their value and the extent to which a good review can, for certain purposes, supersede a large number of primary publications.

SUGGESTED WORKING PAPERS

PRESENT KNOWLEDGE OF INFORMATION REQUIREMENTS

1. Summary of information requirements and problems of scientists in the various fields. This summary might be based on individual statements to be contributed by representative scientists in each field or, alternatively, on informal surveys to be conducted with the aid of interested scientific societies.
2. Summary of findings to date on use of information by scientists.
3. Results obtained through the provision of highly skilled literature scientists as participants in research and deductions concerning information needs.

METHODOLOGY

1. Interviews and surveys to determine information practices and attitudes.
2. The diary method and other types of record keeping on uses of information.
3. Analysis of reference questions and “unanswered questions” to help determine information needs of scientists.
4. Case studies and other methods of determining the role of recorded knowledge in research.
5. Operations research in the area of scientific communication.

Study on the Use of Scientific Literature and Reference Services by Scandinavian Scientists and Engineers Engaged in Research and Development

ELIN TÖRNUDD

EARLIER STUDIES IN THE FIELD

During the past three decades several studies have been carried out to determine the characteristics of the subject literatures used by scientists and the relative importance of different kinds of publications, especially journals in various fields.

These earliest studies as well as the great majority of later ones were based on analysis of library records, tabulation of published material covered by abstract journals and bibliographies, or on reference counting, that is tabulating footnotes and other literature references in journals, dissertations, and books. These studies have thrown light on such questions as trends in publication, the relative importance of different special fields, title dispersion, subject scattering, time span or period of usefulness, language distribution of pertinent literature, and the national origin of materials. Several reference counting studies have produced lists of periodicals most frequently cited in the literature to aid libraries in the selection of periodicals in different subjects. Secondary sources such as handbooks, textbooks, abstracts, and indexes very seldom enter the picture in these studies, because they are not normally referred to among literature references, however often they might have been consulted by the author.

The validity of the results of the above mentioned studies depends first of all on the sample used for tabulation. Regarding tabulation of material in bibliographical

ELIN TÖRNUDD Secretary of the Scandinavian Council for Applied Research.
This study was made under a contract from Unesco.

tools, the results need not indicate its usefulness. References in journals and books need not cover the most useful material consulted, and the material cited need not actually have been used.

E.Brodman (reference 2 below) correlated the relative value placed on specific journals by specialists with the value of these journals as measured through reference counting and came to the conclusion that the latter method is untrustworthy.

Not until recently have studies on the use of literature and other means of communication, directed to the working scientist rather than to his products, entered the scene. Four different methods have been used:

1. Questionnaires directed to *(a)* samples of scientists or *(b)* patrons of libraries ("current use"). This method is criticised for its subjectiveness and the poor correspondence between the recollections by the respondents and the facts obtained from more direct observation.
2. Interviews. This method has the same weaknesses as the previous one, and, in addition, the answers may be influenced by the personal contact with the investigator.
3. Diary. Two of the most competent studies have been carried out with this method, the latest one by R.R.Shaw, who came to the conclusion that the method may be equally unreliable as the two mentioned above. Full reporting is difficult to obtain even under controlled conditions of study.
4. Case study. Because of its slowness and high cost this very valid method requiring an highly trained investigator has been used to a rather small extent. A variation "hit-and-run" method involves short data-gathering visits with the group under investigation, and is less costly than the case study, since less qualified investigators can be used. One of the weaknesses of these methods is that persons under study are aware of being observed.

Large scale programs of operational research studies involving combinations of the above mentioned methods are under way, and it is to be hoped that reliable methods will emerge that will facilitate the production of objective data to build a firm basis for developing the scientific literature and reference services.

Excellent surveys of the literature on earlier studies exist, and among these the following deserve to be pointed out:

EGAN, M., and HENKLE, H.H. Ways and Means in Which Research Workers, Executives and Others Use Information. In *Documentation in Action.* Reinhold Publishing Corp., 1956. pp. 137–159 (bibliography of 53 references).

SHAW, R.R. Studies on the Use of Literature in Science and Technology. In *Pilot Study on the Use of Scientific Literature by Scientists.* National Science Foundation, Washington, 1956. (Informative abstracts of studies published before the beginning of 1954.)

STEVENS, R.E. Characteristics of Subject Literatures. ACRL Monographs. Nos. 5–7, pp. 10–21, Jan. 1953.

A bibliography of earlier studies is presented below. To facilitate a subject approach the latter part of the bibliography is arranged by the field of science or technology with cross references to the first part covering studies carried out with the direct method.

I. BIBLIOGRAPHY OF THE USE OF SCIENTIFIC LITERATURE AND REFERENCE SERVICES AS REVEALED BY STUDIES DIRECTED TO THE SCIENTISTS

1. BERNAL, J.D. Preliminary Analysis of Pilot Questionnaire on the Use of Scientific Literature. In *The Royal Society Scientific Information Conference,* 1948. Report. pp. 101–102, 589–637.

Purpose. To find out what working scientists read, why they read it and what use they make of the information.

Method. Combination of a diary and questionnaire with responses from 208 scientists who represented 8 British research institutes affiliated with the government, universities, and private enterprises.

Results. Tabulated by institutional affiliation, field of science, and status.

This “classic” among studies proved that it was possible to obtain information on current needs of scientists. The revealed data included the following: 37% of sources of references to the literature were references in articles, abstracts 18%, and personal recommendations 14%. Average number of journals consulted varied from 5 to 10 per week. Mean time reported for reading was 5 hours/week. 65% of the sample kept personal indexes, 76% used reviews and all used abstracts. 47% did not read any foreign language easily.

2. BRODMAN, E. Choosing Physiology Journals. *Medical Library Association, Bulletin 32*:479–483, 1944.

Purpose. To check the basic assumptions of the reference counting method.

Method. Physiologists of the faculty of Columbia University were asked to list the periodicals they considered most valuable in order of their usefulness and this list was compared against a list obtained by counting references in an annual review publication and in 3 journals.

Results. The rank correlation between the scientists' list and those obtained through reference counting was low, the corresponding correlation between reference counts from annual reviews and the journals was also low. The reference counting method appeared untrustworthy.

3. BUSH, G.C., GALLIHER, H.P., and MORSE, P.M. Attendance and Use of the Science Library at Massachusetts Institute of Technology. *American Documentation 7:*87–109, 1956.

Purpose. To determine the extent and ways of use of the MIT Science Library.

Method. Questionnaire to 50% of the library users of one week. The 2800 filled in questionnaires were handled by operational research methods.

Results. The users were found to fall into 2 groups: (1) undergraduates who used the library primarily as a study hall staying on an average for 50 min. per visit and (2) graduates who came to use the library materials and stayed about 100 min. per visit. Chemists were observed to use serial literature more than 3 times as much as books while mathematicians used books more frequently. A mathematical model employing probability theory to measure rate and kind of use of library material together with length of stay of the patrons was an outcome of the survey.

4. CURTIS, G.A. A Statistical Survey of the Services of the John Crerar Library. M.A. Thesis, University of Chicago, 1951.

Purpose. To study the extent and way of use of the JCL.

Method. Call slips for materials in the stacks were elaborated with inquiries regarding the intended type of use (school work, private research, company research, etc.) for the publication requested and all patrons were registered. A random sample of the call slips were analysed.

Results. Tabulated by institutional affiliation and use made of the publication, address of the customer, etc. Detailed library statistics emerged.

5. DENNIS, W., and GIRDEN, E. Do Psychologists Read? *American Psychologist 8:*197–199, 1953.

Purpose. To study the reader audience of different features in *Psychological Bulletin.*

Method. Questionnaire to the members of the APA with a 50% response.

Results. The reader audience of different features varied from 9% to 58%.

6. GLASS, B. *Survey of Biological Abstracting.* Johns Hopkins Press, Baltimore, 1954.

Purpose. (1) To study the effectiveness of *Biological Abstracts* and (2) to determine the consensus about it.

Method. For (2) questionnaires to 6995 biologists with responses from 1854. To eliminate the possible bias of the low response rate local samples were interviewed.

Results. Tabulated by the samples: US General, US local samples, US librarians, total US, foreign. The questionnaire results that breadth of coverage was considered satisfactory while the slowness in abstracting was unsatisfactory did not agree with the results of the objective part (1) of the study and appear to indicate that promptness is considered more important than exhaustiveness.

7. GRAY, D.E. Physics Abstracting. *American Journal of Physics 18:*417–424, 1950.

Purpose. To study (1) for what purposes United States physicists use abstracts of physics literature, (2) what they think of the abstracting services available to them, and (3) what they want in the way of abstracts.

Method. A total of 2128 physicists and 300 librarians were queried. A total of 1477 questionnaires from physicists and 202 from librarians were completed. The data were analyzed as follows: (1) by the entire group, (2) in four age groups, (3) in 19 subject subdivisions, and (4) by type of organization of employment.

Results. The seven most used abstracting publications among the physicists were: *Physics Abstracts* (93%), *Chemical Abstracts* (40%), *Nuclear Science Abstracts* (28%), *Electrical Engineering Abstracts* (18%), *Mathematical Reviews* (10%), *Applied Mechanics Reviews* (7%), and *Engineering Index* (7%).

46% of the group used abstracts "principally as a guide"; 6% as substitutes for original articles; 48% half as a guide and half as a substitute. 22% used abstracts principally for keeping up with literature; 30% used them principally for retrospective searches; and the remaining 48% used them for both purposes. In rating *Physics Abstracts,* 96% of the respondents were satisfied with the quality of its abstracts.

86% of the respondents preferred abstracts to mere titles and references. There was a fairly low ceiling on the additional costs that the respondents were willing to pay for an improved abstracting service.

(Abstract condensed from Shaw.)

8. HERNER, S. Information Gathering Habits of Workers in Pure and Applied Science. *Industrial and Engineering Chemistry 46*:228–236, 1954.

Purpose. To determine the sources of research information and reference services that are most useful.

Method. 606 scientists and engineers engaged in research at the various divisions of the Johns Hopkins University were interviewed with the aid of a detailed questionnaire.

Results. Tabulated by nature of research: pure or applied and in some cases by subject field of respondent. The degree of dependence upon scientific literatures as opposed to verbal sources varied and was greater by pure scientists, median 75%, than by applied ones, 50%. The same relation existed between scientists affiliated with academic institutions and non-teaching institutions. The most frequently used direct sources of information were advanced textbooks and monographs, research journals, handbooks, mathematical and physical tables, and research reports. The amount of information derived from domestic journals was 75%. Of the indirect sources of information personal recommendation, references in books and papers, regular perusing of the literature, indexes and abstracts, and bibliographies were considered most useful, and were listed in that order. The average number of journals subscribed to was 2. The reference services furnished by the library were used more than twice as much by applied as by pure scientists. The order of the different services was: accession and selected reading lists, guidance by library staff, bibliographies made on request and translations. 219 scientists were asked whether the literature in its present form is meeting their requirements; 129 were satisfied, 49 had no opinion, and 41 were dissatisfied.

8a. HERNER, S. Library Services. *Chemical and Engineering News 32*:4980, 1954.

Purpose. To determine needs for library service by the technical staff of Atlantic Research Corp. working on jet and rocket fuels.

Method. Interview with research workers.

Results. Distribution of accessions lists and a library bulletin as well as literature searches were in heaviest demand.

9. HERTZ, D.B., and RUBENSTEIN, A.H. *Team Research*. Eastern Technical Publications, New York 1953, 103 pp. (out of print).

Method. The specific areas covered in the study were: (1) the extent of use of the research team, its size, and membership, in various types of research; (2) the organizational relationship and patterns of specific research groups; and (3) communication problems arising from team research activities.

The first phase was a questionnaire survey of the number and kinds of research personnel employed in industrial laboratories. A questionnaire was sent to 3500 industrial research laboratories. 1436 usable responses were received. The responding firms employed a total of 44,639 professional researchers and 21,816 non-professional assistants.

The second phase dealt with the personnel makeup of research teams, and with the extent of "multidisciplinary" group research.

The third phase consisted of field investigations of nine selected research and development groups in five organizations, and of interviews with research administrators and research workers in some 40 laboratories. The questions asked had to do with *(a)* the media and channels of information available to the group and *(b)* the use that is made of these media or channels by the research team or group. The procedures used were:

1. a stratified random sample was asked to fill out a sheet at intervals over a 5-week period telling exactly what he had been doing during the preceding 15 minutes, whether he had communicated with anyone, and whether he had obtained any useful information during this period.
2. a questionnaire which requested information on the communication media each person used for different purposes, the people he contacted most frequently, the number of times he engaged in certain activities over various periods of time, and his best sources of information;
3. personal interviews with a cross-section of the research groups to obtain personal evaluations of available vehicles of communication.

Results. The average number of communicative acts performed by the subjects was two per hour. In 60% of these communicative acts the subject reported a transfer of information. Communication was greatest on Wednesdays, and lower at the beginning of the week than at the end of the week. The proportion of information-bearing communications was highest on Tuesday and Wednesday. The greatest amount of communication took place during the period just before lunch and during the last hour of the day. The highest period for information-transfer was from 10:00 to 3:00.

The project reported was conducted by a staff of 8 full-time members and 17 part-time members over a 2-year period. Two kinds of action are described as basic in the administration of a research or development group. These are *design* and *control*. The basic design-control factors are: physical facilities, equipment, communication media, and personal policies and incentives.

In addition to a mechanism of communication, there must exist (1) the need or occasion for information transfer, (2) a social or psychological rapport among the communicators, and (3) a commonly understood language which permits the ready transmission of intelligence.

As for ways in which information can be transferred, the authors suggest the following: *face-to-face communication* (conversation and conferences), the most used mechanism in the laboratories studied; *written* communication, the least used means of communication: *communication through published material,* least used by organizations in product or process development work and most used among organizations doing fundamental research in physical sciences; *telephone or other mechanical means of* *communication,* an important medium, often neglected by research administrators; *professional* and *society meetings; advanced academic work; visits to outside research organizations.* Operational examinations by the research administrator of all the foregoing means of conveying information are recommended. Propinquity is described as a factor which improves the use of *published material.* On the question (on the questionnaire) as to how industrial research personnel obtain information about how to do their jobs, the first choice of the largest number of individuals was by contacting persons of higher rank. Handbooks and other reference materials kept at their desk were second; contacting persons on the same level, third. Few of the respondents consulted persons of lower rank. A majority considered the library an important source of this type of information, but none considered it first in importance.

Measurements of *communication levels* based on the probability of an individual's giving or receiving information during 29 sample periods revealed the following: the probability among managerial groups was .095; among electromechanical group members, .080; and among design group members, .075. The average probability of an individual either giving or receiving information was .083.

Average communication rates (the average number of contacts in 36 random 15-minute periods) were measured. Based on kinds of people, the rates were as follows: supervisors, .85; professionals .68; assistants, .63. By kinds of jobs, among those on long-term projects the rate was .71, and among those involved in many short-term projects the rate was .68. In working groups of 2 to 4 persons the rate was .35; in groups of 6 to 8 persons the rate was .91; in groups of 9 to 11 the rate was .63.

(Abstract condensed from Shaw.)

10. JOHNS HOPKINS UNIVERSITY, WELCH MEDICAL LIBRARY. *Analysis of Interviews in Indexing of Medical Literature.* Baltimore, 1950. 52pp. (out of print).

Purpose. To study the relative amount of use made of various bibliographical tools, the information gathering habits, opinions on the state of medical indexing services and needs for bibliographic services.

Method. Interviews of 88 medical scientists and 40 medical librarians.

Results. The interviewees seldom used bibliographical tools, and there seemed to be a need for training in the use of available tools. The most useful abstract and index journals were: *Quarterly Cumulative Index Medicus, Chemical Abstracts, Surgeon-General's Index-Catalogue* and *Current List of Medical Literature.* There was no clear preference of indexes over abstracts or vice versa.

11. KENT, A. The Dollars and Cents Value of Company Libraries. Paper delivered before the Executive Conference on Organizing Information. Chicago, Feb. 1, 1957.

Purpose. To study the correlation between company's earning records and library services as well as use of literature.

Method. Questionnaire to 100 companies in the metals field. 45 responses.

Results. In 80% of the 25 top-earning companies, scientific and technical employees spent 6–15% of their working time with literature and 80% of the companies that stated that a great proportion of their scientific people were dissatisfied with the library and its services were in the top 50% in earning rank.

12. KITTEL, D.A. Case Study in Bibliographic Methods. M.A. Thesis. University of Chicago, 1953.

Purpose. To analyze the relative effectiveness of bibliographic devices in social science research.

Method. Case study as member of research team.

Results. Suggestions by those familiar with the literature proved most fruitful.

13. MORRISON, R.A. Use of Current Sources of Information. *I L Research 2*:6–8 (2), 1956.

Purpose. To find out what information sources were used to keep abreast of advances in the industrial and labor relations field.

Method. Questionnaire to 300 firms.

Results. Commercial services, periodicals, reports, conferences, and membership in trade associations were used but not exploited adequately.

14. *Reading Patterns of Engineers in Industry. EPA Technical Information 1*:21–23 (13/ 14), 1955.

Purpose. To study the amount of reading by engineers.

Method. Interview of 200 engineers employed by industries in the Cleveland area.

Results. The average time devoted to reading was 4.5 hours per week with a higher average among subordinates than chiefs.

15. RUBENSTEIN, A.H. Research Communications. *Industrial Laboratories 3*:49–53 (10), 1952.

Preliminary report on investigation. See ref. 9 above.

16. SCATES, D.E., and YEOMANS, A.V. *Activities of Employed Scientists and Engineers for Keeping Currently Informed in Their Fields of Work.* American Council of Education, Washington, D.C., 1950. 35 pp.

Purpose. To find out what working scientists do to keep abreast of developments in their fields.

Method. A total of 1,661 persons were studied. Of this number, 1,087 were in the New York and Philadelphia Naval Shipyards, 46 were in the Bureau of Ordnance in Washington, D.C., and 528 were in industrial firms in metropolitan Philadelphia.

Questionnaires were used in the two Naval Shipyards and in the industrial firms. Since the questionnaires used in the several organizations studied were not the same, the data obtained in each case are not entirely comparable. They are given separately in most instances, and generalizations are made in only a few cases.

Results. Of seven possible self-educational activities (excluding the use of literature, libraries, etc.), the scientists of the Philadelphia Naval Shipyard used the following most frequently: (1) attendance at professional society meetings; (2) attendance at lectures; and (3) attendance at technical conferences. The general participation in self-educational activities was very low.

The scientists of the Bureau of Ordnance showed a very small amount of self-educational activity. The questions put to them included those having to do with the use of professional literature. These questions revealed a relatively small use of the literature. However, 35 different journals were read by one or more persons in the group.

In answer to six questions put to them on a questionnaire, 75% of the scientists and engineers of the New York Naval Shipyard stated that they did a significant amount of professional reading at work and after working hours; 30% attended professional meetings five or more times a year; 27% took occasional courses; 11% wrote technical articles whose preparation required considerable study; and 12% had no time for educational activities of any kind. The amount of activity among the industrial scientists was greater than among those of the New York Naval Shipyard.

Those persons who answered a question on the extent of their reading of periodicals stated that they read an average of seven and a half articles a month. The average number of books read a month was five.

Attendance at professional society meetings and use of other devices for professional contacts were generally low. With the same approximate scoring method for professional contacts and professional reading, the average score for the former was 1.3 while the average score for professional reading was 3.4.

The average number of papers published by the special group was 0.1 a year. The number of contributions was considerably greater among the industrial scientists than among the government scientists.

Among the industrial scientists, there was an increase with age in the amount of professional reading done outside of work. There was an increase in activity as the educational activity increased. There was also an increase with a rise in civil service grade.

Among the various subject fields, the chemists were the most active in self-education, the physicists were second, the electrical engineers were third, the chemical engineers fourth, and the mechanical engineers fifth.

(Abstract condensed from Shaw.)

17. SHAW, R.R. *Pilot Study on the Use of Scientific Literature by Scientists*. National Science Foundation, Washington, 1956. 103 pp.

Purpose. To study the professional reading of the research staff at the Forest Products Research Laboratory, Madison, Wisconsin.

Method. Diary supplemented by a questionnaire kept during two test periods, the first for two months and the second for one month and check of the results through library records, etc.

Results. The average time spent on reading was about 2 hours a week according to the diary information and about 4–5 hours a week according to the estimate of the respondents. Periodicals accounted for 75% of the items read and reports for 14%. 84% of the material was less than a year old. Only 2.5% of the items were in foreign languages.

Only about 43% of the reading acts was recorded, and the diary method used did not prove trustworthy, at least not when administered for long test periods.

18. THE SOCIAL SURVEY. Technical Information in Industry. An International Study of the Dissemination of Technical and Scientific Information to Small and Medium Sized Industry, Carried out for the European Productivity Agency. Mimeographed report, 1957.

Purpose. To survey the methods used by the industries to obtain technical information and thus assist the European information services in planning their programs of action. The study was carried out in Austria, Belgium, Germany, Italy, Norway, the UK, and USA.

Method. Interviewing the owner, manager or the chief technical officer of altogether 2197 establishments 93% of which had 10–500 employees.

Results. Tabulated by country and by size of the establishments: more or less than 100 employees as well as by country and branch of industry (textiles, metals, electrical and food industries). The overall picture that emerged showed that scientific and technical data seldom are used or needed. Only a few establishments had a planned system for obtaining technical information.

	Aus.	*Belg.*	*Ger.*	*Italy*	*Norway*	*UK*	*USA*
The % of establishments subscribing to less than 6 journals	79	62	44	84	36	43	35
The % routing the journals	86	71	78	35	56	74	82
The % of establishments where bulletins or abstracts were prepared regularly or occasionally	8	10	12	2	19	30	34
The % of establishments having 100 or more books	9	5	18	6	25	18	25

Applicable information had been found in the journals by two-thirds of the establishments. Low level of organization=filling 0–2 of the following criteria: (1) having technical advisers, (2) taking more than 5 journals, (3) journals routed according to content and not all journals to all, (4) subscription paid for employees' journals, (5) abstracts prepared, (6) more than 25 books owned, (7) last book added in 1955/1956:

	Aus.	*Belg.*	*Ger.*	*Italy*	*Norway*	*UK*	*USA*
% of establishments with low level organization	56	54	35	79	36	45	34
% of establishments taking foreign language journals	27	78	29	42	82	13	12

The most frequently used method for solving major problems was personal advice in all countries save Belgium and the USA where special research headed the list. About 50% of the respondents stated that published literature was used for the purpose (in Austria and Belgium 23 and 39% and in Germany 66%).

The outside sources of assistance most frequently used were suppliers, consultants, and organizations for fee-paying members such as trade associations.

The main sources of information used by the technical advisers were considered by the respondents to be literature, "inside" know-how and suppliers.

For currently keeping abreast of advances, trade papers, technical and scientific journals, and suppliers formed the most important sources in all countries.

70–90% of the establishments had *not* consulted a library in the last year. In USA, UK, and Norway the libraries consulted were most frequently public reference libraries; elsewhere they were trade association or university libraries.

Recommendations. To follow up the study by investigating the role of suppliers and the reasons why they are successful in imparting technical information and to stimulate the use of existing technical information services on the management level.

19. *A Survey of Research Potential and Training in the Mathematical Sciences.* Final Report of the Committee on the Survey. University of Chicago, 1957. 163+78 pp.

Purpose. To study by what means the mathematical strength of the USA can be increased and under specific items a.o. how the problem of support of mathematical publications may be solved.

Method. (1) Chairmen of departments of mathematics of 61 academic institutions were interviewed. (2) All American and Canadian Ph.D's in mathematics were asked to fill in a detailed questionnaire of 15 pages. Of 2710 mathematicians 68.5% (1851) answered.

Results. (1) Quality of departmental libraries was as a rule considered very good. The book budget varied from $100 to $4000 a year, and the number of current periodical subscriptions from 20 to 310. (2) Tabulation by productivity of the mathematician as revealed by his publishing record: top 15% averaged 1.1–2.3 publications a year (depending on the time elapsed since the Ph.D. degree received), next 35%, 0.3–1 publications a year, next 17%, at least 2 papers during whole career, and the bottom 33%, 1 paper or less during the whole career.

About mathematical publications the present refereeing system was considered satisfactory.

Research articles, expository papers, and reviews were regularly or occasionally read to about the same extent, while the use of abstracts was about 20% less frequent.

One-seventh of the respondents read Russian, but 395 scientists had suffered through inaccessibility of Russian literature and 332 through the inability to read Russian. Only 67 respondents had suffered because of inability to read another foreign language.

299 respondents stated that long delays in publication of the work of others had hampered their research.

615 respondents voted for regrouping journals so that each would specialize but the Committee of the Survey did not consider the action desirable.

Evident need of establishing expository journals was revealed.

20. THORNE, R.G. A Survey of the Reading Habits of the Scientific and Technical Staff at the Royal Aircraft Establishment. Farnborough, 1954. 8 pp. (mimeographed report).

Purpose. To find out how much time was spent by the staff in reading, writing reports, etc., and what type of literature was read.

Method. Diary with supplementing questionnaire was distributed to 300 persons and 91 replies were received with diaries recorded for a period of one week.

Results. Time spent on reading averaged 5.1 hours/week, that on searching and processing literature 2 hours, and that on writing papers 5 hours. Of the items read—average 15.2 per person per week—only 12% were considered of little or no value.

21. TÖRNUDD, E. Professional Reading Habits of Scientists Engaged in Research as Revealed by an Analysis of 130 Questionnaires. M.S. Thesis, Carnegie Institute of Technology, Pittsburgh, 1953.

Purpose. To study the information gathering habits of researchers at the Mellon Institute.

Method. Questionnaire to 136 scientists, response rate 95.6%.

Results. Average weekly reading time 5.2 hours. Number of journals personally subscribed to 4.1 and regularly read 7. Personal indexes kept by 2/3 of the sample.

Translations from foreign languages needed by *ca.* 50%. The need to learn more about the use of subject literatures had been felt by 44%.

The most important sources of information in their order of usefulness were: journals, books, meetings, abstract journals.

22. UNIVERSITY OF MICHIGAN, Institute for Social Research, Survey Research Center. *The Attitudes and Activities of Physiologists.* A nationwide study. The University, Ann Arbor, 1954, 155+187 pp.

Purpose. To study by what means the physiological field can be promoted and the impact of physiology on the national welfare. Under specific aims a.o. to collect and assess information concerning communications.

Method. A detailed 10-page questionnaire was sent to 7104 physiologists—5393 responses.

Results. Tabulated by different fields of physiology and "central"=mainly physiologist as opposed to "peripheral"=using physiology as auxiliary science.

The publishing record for the period of 3 years averaged 4–5 papers (26% 0–2 papers, 30% 3–6 papers, 16% 7–10 papers, and 15% 11 or more papers). There was a curvilinear relationship between the number of publications and age. 5 papers and more were published by the following percentages in different age groups: under 30 years 29%, 30–39 47%, 40–49 60%, 50–59 54%, 60–69 56%, and 70– 34%. A direct relationship was found between publishing record and the salary, research funds, the ease of obtaining research funds, and the freedom of choosing his research problem. The publishing record is higher among academic and governmental physiologists than among industrial ones.

The relative importance of different sources of information is indicated by the % of respondents using each considerably or to some extent (figures in parentheses refer to considerable use): journals 98% (86), abstracts 82 (41), reviews 77 (31), monographs 64 (20), meetings 72 (25) and conferences 62 (21).

59% of respondents considered journals satisfactory as they are at present, 48% abstract journals, 48% reviews, 40% monographs, 30% meetings, and 28% conferences.

The ability to keep up with advances in the field was considered very good by 11%, fairly good by 66%, not too good by 20%, and not good at all by 2%.

Problems that hampered the ability to keep abreast of advances were time limitation indicated by 48% of the respondents, too many publications and too large field by 30%, lack of access to published material by 11%, slowness of publications by 8%, inadequacy of publications by 7%, inadequacy of abstracts, indexes and reviews by 6% and isolation from colleagues by 6%.

23. URQUHART, D.J. The Distribution and Use of Scientific and Technical Information. The Royal Society Scientific Information Conference, 21 June–2 July, 1948, Report, pp. 408–419. The Royal Society, London, 1948.

Purpose. To determine: *(a)* how references to materials requested were obtained, *(b)* what the required information was needed for, and *(c)* whether the requested publications contained the desired information.

Method. The group surveyed consisted of borrowers of the Science Museum Library in London. Short questionnaires were enclosed in the requested publications of 715 borrowers over a 2-week period in 1949.

Of the 715 questionnaires dispatched, 354 were returned. A combination of simple tabulations and Batten punched-cards was used in the reduction of the data obtained.

Results. (1) Abstracts were main source of citations, followed closely by references in periodical articles. (N.B. The residual nature of materials borrowed from SML.)

(2) About half the references taken from abstract journals were for the past 12 months. This emphasizes that abstracts were used heavily by scanning them as received.

(3) Fifty percent of the material called for was US material (41% for 1931–40, 52% 1941–45, and 67% for 1947).

(4) One-fourth of the literature consulted was published in the last 1 1/2 years, and over 50% in the last 6 years.

(5) In 77% of the cases studied, the publication requested contained the required information.

(Abstract condensed from Shaw.)

24. URQUHART, D.J. Public Libraries and Industry. *Manchester Review 6*:468–472, 1953.

Purpose. To determine how industry obtains technical information.

Method. Interviews.

Results. (1) Less than 2% of the manufacturing establishments in the United Kingdom maintain libraries; 69% of all firms were found to have less than five periodicals of any sort.

(2) The vast majority of people in industry were found to have no knowledge of how to obtain technical literature and technical information. In the majority of the firms studied there was nobody assigned to obtain and circulate technical information.

(Abstract condensed from Shaw.)

25. WAPLES, D. Belgian Scholars and Their Libraries. *Library Quarterly 10*:231–243, 1940.

Purpose. To determine which libraries Belgian university professors used for obtaining material in different fields.

Method. Questionnaire.

26. WILLIAMS, G.R. A Study of the Bibliographic Sources Used by the Patrons of the John Crerar Library. M.A. Thesis, University of Chicago, 1952.

Purpose. To determine the relative extent of use of the various parts of the bibliographic machinery in JCL.

Method. Detailed questionnaire to library patrons who borrowed material held in the stacks (current journals and handbooks, etc., are shelved in the reading rooms).

Results. Footnotes in articles and other references were found to be relied upon to a great extent—footnotes to a greater extent than references in bibliographies following publications. (N.B. the study did not cover current periodicals.)

68.8% of the respondents had had training in the use of libraries and 36.6% had received training in the literature of their field.

27. ÖHMAN, E. Jernkontorets Litteraturöversikt. *Tidskrift för dokumentation* 5:67–70, 1949.

Purpose. To determine whether Jernkontorets Litteraturöversikt, the selective abstract journal in the field of iron and steel published by the Swedish Iron Mongers' Research Association, was useful to its readers.

Method. A questionnaire to the subscribers with 110 responses.

Results. The preference for fullest possible informative abstracts 3–6 months after appearance of the original article as opposed to prompt indicative abstracts was indicated by a great majority.

30% of the respondents filed clippings from the bulletin in their personal indexes.

Only 60% of the respondents had convenient access to foreign abstract publications in the field and only 40% made regular use of them.

27a. Moss, L., and WILKINS, L.T. Studies in the Use of Technical Information in the Smaller Industrial Establishments. Paper 16, presented at a Symposium on The Direction of Research Establishments, National Physical Laboratory, Sept. 1956. 20+10 pp.

Purpose. The Social Survey studied the use of information by individual scientists and technologists employed in 127 British electrical firms with 200 to 1000 employees.

Method. Interviews with 1082 persons of whom under 20% had some kind of university training.

Results. The preliminary report presents data on the nature of duties analyzed by qualifications and by experience. Those concerned with research and development were found to be on average more highly qualified than others, and 75% of the group concerned with production supervision and control had no qualifications. The size of firm did not seem to be associated with information collecting methods. Whilst literature was recognized as the most useful source of new ideas, in practice only a minority went to literature when working on current problems. The use of literature as an aid to problem-solving was clearly related to the level of qualification. Technical journals were found to play a much more important role than scientific ones, abstracts, and reprints. Mass communication media like newspapers, radio, and TV were rated low as information sources. Less than 10% of the firms had a specific official whose function it was to draw the attention of scientific and technical staff to useful information.

II. BIBLIOGRAPHY OF EARLIER STUDIES ON THE USE OF SCIENTIFIC LITERATURE AND REFERENCE SERVICES ARRANGED BY SUBJECT FIELD (with references to the section above)

UDC 1 Psychology See also references 5 and 8 above and 69 below.

28. BROADUS, R.N. The Research Literature of the Field of Speech. A.C.R.L. Monograph 7:22–31. Association of College and Reference Libraries, Chicago, 1953.

Purpose. What forms of publications are used? From what other fields are information and techniques gathered? In what languages? What ages of publications are used?

Method. Reference count in four journals covering various sample periods from 1919 to 1951.

Included newspapers, books, and periodicals.

Results.

1. The average number of footnotes per thousand words of text has increased threefold over the years.
2. In various parts of the field serials made up from 34 to 62% of the references and books ranged from 32 to 63%.
3. Classification of the subject matter of the 6700 citations showed that it was widely distributed over all fields of knowledge.
4. The languages averaged as follows:

   English 92%, German 4.1%, French 1.2%.
5. The average ages of publications used showed much less concentration on current years than is the case in science and technology, yet 27% was in the last five years and over 60% in the last 20.

(Abstract condensed from Shaw.)

29. DANIEL, R.S. and LOUTTIT, C.M. A Survey of Psychological Literature. In *Professional Problems in Psychology,* pp. 35–66. Prentice-Hall, New York, 1953.

Purpose. To study the scatter of psychological publications and dispersion of titles.

Method. Bradford's “Law of Scattering” (see 55 below) is applied to the measurement of the dispersion of the literatures on psychology, chemistry, and physics (using data by Fussler, see 42 below).

Results. Very few publications of interest to psychologists appear in psychology journals.

30. LOUTTIT, C.M. The Use of Foreign Languages by Psychologists. *American Journal of Psychology 68*:684–6, 1955.

Purpose. To determine whether psychologists draw on the world literature and cite foreign references proportional to the volume of foreign literature in the field.

Method. Reference counting in 7 journals.

Result. Hypothesis not supported.

31. LOUTTIT, C.M. The Use of Foreign Languages by Psychologists, Chemists and Physicists. *American Journal of Psychology* 70:314–6, 1957.

Purpose. To test the validity of the hypothesis of proportional citation of foreign references to the distribution of material by country and language.

Method. Reference counting.

Results. The English and German, especially psychologists, cite their own language far in excess of its relative proportion while the French rely less on French. French chemists and physicists cite other than their own language in 70% of cases.

32. LOUTTIT, C.M. Publication Trends in Psychology 1894–1954. *American Psychologist 12*:14–21, 1957.

Purpose. To determine trends in subject interest, journal sources in psychological literature, and language.

Method. Reference counting of entries in every 5th volume of *Psychological Index* and *Psychological Abstracts.*

Results. Literature in the applied field has increased. German language shows a decline and English a corresponding increase.

UDC 3 Social Sciences See also references 12 and 13 above.

33. HOBBS, A.H. *The Claims of Sociology, a Critique of Textbooks.* Stackpole Co., Harrisburg, 1951.

Method. Reference counting in books.

34. LIVESAY, M.J. Characteristics of the Literature Used by Authors of Books in the Field of Economics. M.A. Thesis, University of Chicago, 1952.

Method. Reference counting in books.

35. MARTIN, G. Characteristics of the Literature Used by Authors of Books on Political Topics. M.A. Thesis, University of Chicago, 1952.

Method. Reference counting in books.

36. MEIER, E.L. Characteristics of the Literature Used by Contributors to American Sociological Journals. M.A. Thesis, University of Chicago, 1951.

Method. Reference counting in journals.

37. QUINN, E.W. Characteristics of the Literature Used by Authors of Books in the Field of Sociology. M.A. Thesis, University of Chicago, 1951.

Method. Counting references in books.

UDC 51 Mathematics See also references 1, 3, 8, and 19 above.

38. ALLEN, E.S. Periodicals for Mathematicians. *Science* 70:592–594, 1929.

Purpose. To determine which mathematical journals were most used in 1929.

Method. Reference counting in American and foreign journals.

Results. Over 50% of the citations referred to 9 journals. English journals were cited most frequently.

39. BROWN, Ch.H. Scientific Serials: Characteristics and Lists of Most Cited Publications in Mathematics, Physics, Chemistry, Geology, Physiology, Botany, Zoology and Entomology. Association of College and Reference Libraries Monograph No. 16, Chicago, 1956.

Purpose. To determine the relative importance of 838 serials to science as a whole and in different fields of science.

Method. Reference counting.

Result. Recommended lists of important journals.

40. HOPP, R.H. A Study of the Problem of Complete Documentation in Science and Technology. *Dissertation Abstracts 16*:1689*,* 1956.

Purpose. To determine the dispersion of papers in mathematics, chemistry, biology, and physics.

Method. Reference counting in bibliographies.

Results. A considerable portion of literature was concentrated within a few key periodicals.

UDC 53 Physics See also references 3, 7, 8, 16, 17, 23, 29, 31, 39, and 40.

41. HOOKER, R.H. A Study of Scientific Periodicals. *Review of Scientific Instruments 6:*333–338, 1935.

Purpose. To determine the most used journals in physics and radio in 1935.

Method. Reference counting in 5 journals.

Results. A heavy concentration was found in very few journals. The radio list was headed by the most important physics journals.

42. FUSSLER, H.H. Characteristics of the Research Literature Used by Chemists and Physicists in the United States. *Library Quarterly 19*:19–35*,* 119–143, 1949.

Purpose. To determine the research literature used in the US in “pure” chemistry and physics: (1) the importance of literature of various subject fields of chemistry and physics; (2) the temporal span, particularly the span between date of original publication and date at which it is known to have been used; (3) the principal forms of literature used and their relative importance; (4) the national origins of literature used; (5) the important serial titles for each field.

Method. Reference counting in *Physical Review* and *Journal of the American Chemical Society*.

Results.

1. In general, the serial literature used for research in physics is more recent than that in chemistry. In chemistry, around 50% of the citations are within 5 years of the year of reference and around 70% are within 10 years. In physics, around 60 to 70% are in the first 5 years and from about 70 to about 87% are in the last 10 years.
2. A definite shift is shown from the use of foreign literature to the use of American literature.
3. In both chemistry and physics, serial citations make up somewhat better than 90% of all citations. Monographs account for 5 to 6% of the citations in chemistry and 7 to 11% in physics. The use of patents in chemistry has been increasing and as of 1946 accounted for almost 2% of the citations.

(Abstract condensed from Shaw.)

43. SCHAUBER, A. An Analysis of the Documentation of Physics Research to Determine the Serials Most Frequently Used. M.S. Thesis, The Catholic University of America, 1951.

Method. Reference counting in journals.

UDC 54 Chemistry See also references 1, 3, 8, 16, 17, 23, 29, 31, 39, 40, and 42.

44. BARRETT, R.L., and BARRETT, M.A. Journals Most Cited by Chemists and Chemical Engineers. *Journal of Chemical Education 34*:35–38, 1957.

Purpose. To determine the journals read by American chemists.

Method. Reference counting in *Industrial and Engineering Chemistry,* July-Dec. 1955, and *Journal of the American Chemical Society,* July-Aug. 1955.

Results. Relative importance of German has declined since studies made in 1933 and 1939. Much Russian, Japanese, and Italian literature seemed overlooked.

45. CRANE, E.J. Periodical List of Periodicals. *Chemical and Engineering News* 25:2075, 1947.

Purpose. To present a survey of the coverage of *Chemical Abstracts* in 1946.

Method. Tabulation of periodicals abstracted in *CA* in 1936 and 1946.

Results. During the 10 years the number of periodicals had increased by 50%. English was the most frequently abstracted language followed by German, Russian, French, Spanish, Japanese, and Italian.

46. GROSS, P.L.K., and GROSS, E.M. College Libraries and Chemical Education. *Science 66:*385–389, 1927.

Purpose. To determine periodicals used most by chemists in 1926.

Method. Reference counting in the *Journal of the American Chemical Society* for 1916–1925.

Results. This classic first study lead to the conclusion that the relative importance of journals varies with a high speed. Five journals accounted for more than 50% of the total citations, although not the same five at different times. More than 50% of the citations were at that time German.

UDC 55 Geology and Meteorology See also references 1, 8, 39, and 55.

47. BAUM, W.A. A Study of Reference Citations in the *Journal of Meteorology* and the *Quarterly Journal of the Royal Meteorological Society. Bulletin of the American Meteorological Society 36*:61, 1955.

Purpose. To determine the most cited meteorological journals.

Methods. Reference counting in the two journals.

48. GROSS, P.L.K., and WOODFORD, A.O. Serial Literature Used by American Geologists. *Science* 73:660–664, 1931.

Purpose. To determine the periodicals most used by geologists and mineralogists.

Method. Reference counting in 6 American journals for 1927, 1928, and 1929.

Results. The seven most frequently cited journals covered almost 50% of the references, and almost 50% of the cited literature was German in 1929.

UDC 56/59 Biological Sciences See also references 1, 3, 6, 8, 17, 23, 39, 40, and 69.

49. HENKLE, H.H. The Periodical Literature of Biochemistry. *Medical Library Association Bulletin 27*:139–147, 1938.

Purpose. To determine the most used periodicals in biochemistry.

Method. A *reference count* based on the *Annual Review of Biochemistry*. The years 1932 through 1936 were checked.

Results. The distribution of the literature was found to illustrate the law of diminishing returns. Of the 17,198 references to 851 periodicals almost one-half referred to 10 journals. Forty-six percent of the articles cited were less than 5 years old and 59% less than 10 years old.

(Abstract condensed from Shaw.)

50. HINTZ, C.W.E. Internationalism and Scholarship: A Comparative Study of the Research Literature Used by American, British, French, and German Botanists, Thesis, University of Chicago, 1952. 175 pp.

Purpose. To ascertain the extent of interchange of scientific information through the literature among botanists of the United States, Great Britain, France, and Germany.

Method. A *reference count* from arbitrarily selected US, German, and French journals in botany, tabulating in each all domestic citations for the last 5 years.

Both in 1895 and 1939 serial literature was used more often than non-serial literature. Between these 2 years, there was a general increase in the proportionate dependence upon serial literature.

Recency of publications did not appear to be as important a factor to the botanists as to the chemists and physicists.

The extent and manner of use of scientific literature increases as scientific activity increases.

(Abstract condensed from Shaw.)

UDC 61 Medicine See also references 1, 2, 8, 11, 22, and 39.

51. HACKH, I. The Periodicals Useful in the Dental Library. *Medical Library Association Bulletin 25*:109–112, 1936.

Purpose. To determine the most important journals for a dental school library.

Method. Reference counting in about 20 periodicals.

Result. 6 journals were found to supply 50% of the references.

52. HUNT, J.W. Periodicals for the Small Biomedical and Clinical Library. *Library Quarterly 7*:121–140, 1937.

Purpose. To aid in selection of periodicals.

Method. Tabulation of loan records for 1934–1935.

Results. Tabulation of frequency of loan by 5-year periods shows 192 journals circulated 12 or more times.

Circulation by publication date indicates that 52% of the periodicals borrowed were issued in the last 5 years, 74% in 10 years and 85% in 15 years.

(Abstract condensed from Shaw.)

53. JENKINS, R.L. Periodicals for Medical Libraries. *American Medical Association, Journal 97*:608–610, 1931.

Purpose. Selection of periodicals for a small clinical library.

Method. Reference count in three foreign and American journals for 1928–1929; weighting the foreign citations to avoid national bias.

Results. The order of references were: American, English and then German.

Ten periodicals contain 35% and 20 contain almost 50% of the citations.

Eighty-two percent were, in the last 10 years.

(Abstract condensed from Shaw.)

54. SHERWOOD, K.K. Relative Value of Medical Magazines. *Northwest Medicine 31:*273–276, 1932.

Purpose. To determine what books and magazines the doctor should buy.

Method. Reference count of all citations in one year of the *A.M.A. Journal.*

Results. 4186 references to periodicals, 422 to books and 201 to special reports and to society transactions.

55 percent of the references fell within the preceding 5 years and 75 percent in the last 10 years.

Considering the last 5 years only and the whole period the rank of the first four periodicals was the same, and provided 22 percent of the references.

The most frequently cited languages were English, then German, with French third.

(Abstract condensed from Shaw.)

UDC 621 Mechanical Engineering See also references 1, 8, 14, 16, 17, 18, and 23.

55. BRADFORD, S.C. *Documentation,* 2d edition. Crosby Lockwood & Son, London, 1953. Chapter IX.

Purpose. To study the adequacy of coverage of scientific and technical abstract journals and the scatter of articles on applied geophysics and lubrication.

Method. Counting abstracts in 300 abstracting and indexing journals and checking duplicate abstracting as well as counting references in bibliographies.

Results. Only 280,000 useful articles of the estimated total of 750,000 were found to be abstracted, and those abstracted were found to have been abstracted on an average 2.7 times (a check for electrical engineering yielded the duplication factor 3.9). The law of scatter was developed.

56. VOIGT, M.J. Scientific Periodicals As a Basic Requirement for Engineering and Agricultural Research. *College and Research Libraries 8*:354–359, 375, 1947.

Purpose. To study the interrelationship between scientists and research carried out in the applied fields of engineering and agriculture. Particular emphasis was given to the analysis of subject fields in the pure sciences as they relate to the applied sciences.

Method. Reference counting in the subjects soils and dairying in the *agricultural* field and *mechanical* and *metallurgical* engineering in the technical field.

Results. Metallurgical engineering makes a great deal of use of the publications of pure science with 23% of the references in pure science. In mechanical engineering the references used in pure science were slightly less than half this percentage. Three-fourths of the references cited in metallurgical journals were in periodicals in the technical and engineering field, with 61% in mining and metallurgy. In mechanical engineering 87% of the references were distributed in various engineering fields.

In the two fields of agriculture the pure science periodicals accounted for about 31% of the total references in the soils journal and more than 35% of the periodicals cited in the dairying journal. This indicates that the pure science journals are of importance in applied science as well as in pure science.

(Abstract condensed from Shaw.)

UDC 621.3 Electrical Engineering See also references 16, 18, 27a, 41, and 55.

57. COILE, R.C. Periodical Literature for Electrical Engineers. *Journal of Documentation 8*:209–226, 1952.

Purpose. To determine the periodicals used by electrical engineers.

Method. Reference count from six journals for 1949; omitting self references in ranking.

Results. Eighteen journals covered 50%; 54 journals contained 75% of the references. The age of papers cited was: 50% less than 4 years; 75% less than 11 years.

(Abstract condensed from Shaw.)

58. DALZIEL, C.F. Evaluation of Periodicals for Electrical Engineers. *Library Quarterly 7*:354–372, 1937.

Purpose. To determine the most important journals for electrical engineers and to find a practical means of evaluating scientific periodicals.

Method. Reference counting method was selected.

Results. Plotting the number of references to a given periodical as found in the present study on log-log paper, the author found that he obtained a curve that was very similar to that obtained by plotting the results of previous studies. From this, he concluded that reference counts in chemistry, mathematics, physics, radio, and electrical engineering follow a simple mathematical law.

A method for measuring the number of references needed for a valid count is discussed. It consists of tabulating references and plotting them on log-log paper until the first 20 or 30 points determine a straight line. This indicates that a sufficient number of references have been obtained to follow the probability law controlling this type of investigation.

(Abstract condensed from Shaw.)

59. MCNEELEY, J.K., and CROSNO, C.D. Periodicals for Electrical Engineers. *Science 72*:81–84, 1930.

Purpose. To determine deficiencies in periodical holdings at Iowa State College by determining the journals used most frequently by electrical engineers.

Method. Reference count in seven key foreign and domestic journals for January 1925 through June 1929.

Results. Nine most frequently cited journals produced over 50 percent of the citations.

(Abstract condensed from Shaw.)

UDC 624 Civil Engineering

60. RITCHIE, M.G. An analysis of the documentation of civil engineering research to determine the serial publications most frequently used. Thesis, M.S., Washington, D.C, The Catholic University of America, 1951. 58 pp.

Purpose. To determine:

1. What are the important titles in the field?
2. What is their relative importance?
3. What related fields of literature are important?
4. What are the important titles in these related fields?
5. What foreign languages are necessary to civil engineers?
6. What foreign journals are necessary?
7. What is the importance of supplying back volumes? How far back in point of time do civil engineers go for their reference material in research?

Method. The source journals used in the present study were: *Transactions of the American Society of Mechanical Engineers, Journal of the American Water Works Association, Sewage Works Journal, Proceedings of the American Concrete Institute,* and the *Proceedings of the Highway Research Board.* The volumes for the period 1945 through 1949 were analyzed. Duplicate references and references to unpublished materials were not recorded.

Results. Twelve titles account for 50% of the references. The first 100 titles cited accounted for 78% of the references.

(Abstract condensed from Shaw.)

UDC 629.13 Aernautical Engineering See also reference 20.

61. RANDALL, G.E. Who Uses a Technical Library.

Purpose. To study the current use of Arnold Engineering Development Center.

Method. Analysis of loan records.

Results. Only 27 of total staff (74% of the engineers) used the library.

UDC 63 Agriculture See also references 23 and 56.

62. CROFT, K. Periodical Publications and Agricultural Analysis. *Journal of Chemical Education* *18*:315–316, 1941.

Purpose. To determine the chemical magazines used most by agricultural chemists.

Method. Reference count in 23 volumes of the *Journal of the Association of Agricultural Chemists,* listing all periodicals referred to 10 times or more between 1915 and 1940, omitting government publications.

Results. Only 11 journals averaged as many as five references per year over the whole period and six accounted for half the use.

(Abstract condensed from Shaw.)

UDC 66 Chemical Engineering and Technology, Food Technology, Oil Technology, and Metallurgy See also references 8, 8a, 11, 16, 18, 21, 23, 27, 44, and 56.

63. BORG, F.S., and LOFTMAN, K.A. Domestic and Foreign Periodicals in the Field of Petroleum Chemistry. *Oil and Gas Journal* 199–208, April 21, 1949.

Method. Tabulation of references in *Chemical Abstracts.*

64. HARDIE, B.G. A Study of the Use of Pure Science Periodicals in Practical Research on Petroleum. M.S. Thesis, Carnegie Institute of Technology, 1949. Also *Oil and Gas Journal 121,* 159–61, May 18, 1950.

Method. Reference counting in journals.

65. SMITH, M.H. The selection of chemical engineering periodicals in college libraries. *College and Research Libraries 5,* 217–227, 1944.

Purpose. Selection of chemical engineering journals.

Method. Reference count from four journals and one handbook, including self-references.

Results. A total of 21,728 references were counted. The bulk came from *Industrial and Engineering Chemistry*. The number of foreign references in all the American sources was less than 50%. In the British source it was more than 50%. German was by far the most important of the foreign languages; French was second.

(Abstract condensed from Shaw.)

UDC 677/678 Textile and Rubber Technology

66. PURDUM, C.W. A Statistical Analysis of Articles Appearing in the Literature of Rubber Chemistry. M.A. Thesis, Western Reserve University, 1951.

Method. Tabulation of references in *Chemical Abstracts.* For textile industry see reference 18.

UDC 72 Town Planning

67. MORRIS, J.E.B. The Library Materials Used in Urban Planning. Differences Between Items Cited by Scholars and by Practitioners within the same Field. M.A. Thesis, University of Chicago, 1955.

Method. Reference counting in books.

Result. Scholarly works cite most frequently theoretical works while practical ones draw mainly on descriptive literature.

UDC 93 History

68. ALSTON, A.M. Characteristics of Materials Used by a Selected Group of Historians in their Research in United States History.

Method. Reference counting in periodicals and books.

69. STEVENS, R.E. The use of library materials in doctoral research; a study of the effect of differences in research method. *Library Quarterly 23*:33–41, 1953.

Purpose. To prove or disprove the seeming differences between the character of historical research and its literature and that of experimental research and its literature.

Method. Reference counting in representative doctoral dissertations. The thesis collections of three universities, Washington University (St. Louis), the University of Michigan, and the University of Illinois, were studied. The subjects chosen for investigation were the following: United States history, classical language and literature, botany, psychology, and education. The same number of dissertations was chosen in each field. A total of 100 dissertations were examined. They were classified as to research method (historical or experimental) by a team of judges. Upon analysis, it was found that some of the dissertations followed a third method, which was termed “textual.” Textual dissertations were defined as tracing the technical tradition of a manuscript. Of the 100 dissertations, 40 were classified as historical, 9 were textual, and 51 were experimental.

Results. In the historical research dissertations, the mean number of citations of titles not in the university library was 65.67. In textual research the number was 9.33. In experimental research, the number was 3.71. The mean number of titles cited only once in historical research dissertations was 128.85. This number for textual research was 33.89. For experimental research it was 16.76. Thus, it is shown that there are a greater *number* and a greater *proportion* of rarely issued titles, and of titles not in the library, cited in historical than in experimental research.

When often-cited titles were compared with titles cited only once, it was found that the proportion of the more-often-cited titles not in the library was much lower than that of the less-often-cited titles. When the data obtained by considering the 100 theses as a unit were broken down by library, by specific subject field, and by type of research, the fact still emerged that the proportions of titles cited in historical dissertations that were not in the library was greater than in experimental dissertations. Further, the author found by correlating rarely used titles in all fields and types of research with the number of cited titles not in the library that these two facts are closely related.

(Abstract condensed from Shaw.)

THE SCANDINAVIAN STUDY

A study was set up in Denmark and Finland to supplement the results obtained in previous studies (references 1–27a) of which a great majority have been carried out in the USA (22 studies) and in the UK (6), while the European scene has been entered only by the EPA study covering a sample of small and medium sized industries (reference 18) and by two smaller investigations (references 25 and 27).

In order to ascertain the requirements for scientific and technical information of those who are liable to be encountering the greatest problems the study was designed to cover two samples of junior research workers within the age limits 25–40 years excluding professors, directors of institutes, and other persons in top positions, who can be expected to know personally most of the scientists in their own field and be in the position to assign literature work to their assistants. In both countries the samples were drawn from three different working environments: industrial laboratories, research institutes, and academic institutions.

Through the Danish Academy of Technical Sciences (ATV) its Junior Scientists Committee was approached. The Board of this group showed a great interest in the study and supplied the required information of their contact net from which a sample of 100 persons was chosen at random.

The Finnish sample of 100 younger scientists was chosen at random from a register of scientific and technical research workers.

During September-November 1957, 100 copies of the Danish and 100 copies of the Finnish questionnaire together with a circular letter, appended to the end of this report in an English translation, were mailed to the two samples and 190 filled-in questionnaires were received. Among these were two that had to be omitted from the tabulation, as the respondents were overqualified for the groups in question. Negative responses were received from four scientists who recently had moved to another job not involving research activities, or had finished working to become housewives. After two reminders only 6 cases remained unanswered on Dec. 31, 1957, and a response rate of 97% was considered satisfactory.

The answers were coded on standard type key-sort cards and tabulated as a rule by country and institutional affiliation. The latter approach was selected rather than grouping by the nature of research activities because: (1) the distinction between basic and applied research was subject to individual interpretation and both types of activities were in several cases carried out by the same person; (2) only about a fifth of the respondents considered their present work to be primarily of basic nature; (3) the institutional approach was considered most rational for the purpose of planning measures to improve future services.

Among the institutional affiliations the “academic” section comprises scientists and engineers working at universities, institutes of technology, and other academic institutions in which the majority of respondents divide their time between research and teaching. The “research” section employed at research institutes, cooperative institutes sponsored both by branches of industry and by government institutes, comprises the most nearly full-time researchers in the samples, while most of the respondents in the “industrial” section divide their time between research and development on one hand and production supervision on the other.

The results are grouped by questions.

I. THE STUDY POPULATION (QUESTIONS 1–2)

The distribution of respondents by field of research is given in Table 1. The Danish and Finnish groups differ little from each other. Among the latter there were fewer civil, electrical, and mechanical engineers and correspondingly a stronger representation on the chemical engineering side.

TABLE 1 *Distribution of respondents by field of research and institutional affiliation (Questions 1 and 2; based on 188 responses)*

Institution	*Field of research classified by UDC*								
	51/53	*54*	*55/59*	*61*	*62*	*63/65*	*66/664*	*665/669*	*674/678*
Danish (94)									
Academic	8	10	2		4	2		1	
Research institutes			5	1	15	2	3	1	2
Industry	1				7	2	10	14	4
Total Danish, %	10	11	7	1	28	6	14	17	6
Finnish (94)									
Academic	4	8	1	1	2	1	2		
Research institutes	1	1	5	1	8	1	5	3	3
Industry				2	5	3	14	14	9
Total Finnish, %	5	10	6	4	16	5	22	18	13
Both Groups, %	7.5	10.5	6.5	2.5	22.0	5.5	18	17.5	9.5

51/53 mathematics, astronomy, physics }
54 chemistry } "pure scientists"
55/59 geology, meteorology, biology }
61 pharmaceutical technology
62 engineering (excluding chemical engineering)
63/65 agriculture, household equipment, industrial management
66/664 chemical engineering and technology, industrial microbiology and food technology
665/669 oil, ceramic, paint, and soap technology, metallurgy
674/678 wood, rubber, paper, textile, and plastics technology

As seen from Table 2 the study population consisted of 126 graduates from institutes of technology and 62 graduates from universities, agricultural colleges, and pharmaceutical colleges. All respondents had an academic record corresponding to an American M.S. degree. In addition 12 respondents in the group of engineers and 11 respondents in the group of scientists had doctoral degrees.

The academic section of 27 persons in the Danish sample was somewhat larger than that in the Finnish one. In both groups the research institute sections were of the same size, while the industrial section had a heavier representation in the Finnish group. The 41 respondents who considered their work to be primarily basic research were all affiliated with academic institutions save 9 who worked in research institutes.

TABLE 2 Distribution of respondents by academic degree, nature of research and institutional affiliation (Question 2; based on 188 responses)

	Doctor of Tech.			*M.S. Engineering*			*Ph.D. Science*			*M.S. Science*			*All*	
Institution	*Pure*	*Appl.*	*Both*	*Pure*	*Appl.*	*Both*	*Pure*	*Appl.*	*Both*	*Pure*	*Appl.*	*Both*	*Pure*	*Appl.*
Danish														
Academic	3	1	4	6	5	11	2		2	9	1	10	20	7
Research institutes					20	20		1	1	4	4	8	4	25
Industrial		1	1		33	33					4	4		38
Total Danish	3	2	5	6	58	64	2	1	3	13	9	22	24	70
			(5%)			(68%)			(3%)			(23%)		
Finnish														
Academic	4	2	6	4	3	7	3	1	4	1	1	2	12	7
Research institutes					11	11		3	3	5	9	14	5	23
Industrial		1	1		32	32		1	1		13	13		47
Total Finnish	4	3	7	4	46	50	3	5	8	6	23	29	17	77
			(7%)			(53%)			(9%)			(31%)		
Both groups	7	5	12	10	104	114	5	6	11	19	32	51	41	147
			(6.4%)			(60.6%)			(5.9%)			(27.1%)	(21.8%)	(78.2%)

II. HOW WELL DO YOU FEEL YOU ARE ABLE TO KEEP UP WITH ADVANCES IN YOUR FIELD? (QUESTION 3)

As seen from Table 3 the academic sections appeared to be in the best position in regard to the ability of keeping abreast with current advances. This could be expected considering the nature of work: mostly long-term research projects and teaching in specific subjects. The research sections were least satisfied with their ability to tackle the problem. This finding probably reflects the difference in activities: short-term research projects in a variety of related subjects, and less “academic freedom.” Respondents in the industrial section took the position in between the academic and research sections. The overall Danish picture checks very well with the results of the American study of physiologists (22).

TABLE 3 Estimated ability in keeping up with the new developments[a] (Question 3; based on 188 responses)

	very well		*satisfactorily*		*not well at all*		*Total No*
Institution	*No.*	*%*	*No.*	*%*	*No.*	*%*	
Danish							
Academic	5	18	19	70	3	12	27
Research institutes	2	7	19	66	8	27	29
Industrial	6	16	26	68	6	16	38
Total Danish	13	14	64	68	17	18	94
Finnish							
Academic	10	53	9	47			19
Research institutes	6	21	22	79	28		
Industrial	16	34	31	66	47		
Total Finnish	32	34	62	66	94		
Both groups	45		126		17		188

[a] No correlation was found between estimated success in keeping up with advances and location of institution: in the large cities or at a distance from these centers.

The Danish and Finnish groups cannot be compared with each other, because of an unfortunate mistake in the wording of the given alternatives in the two languages. The Finnish alternatives were (1) well, (2) fairly, (3) badly; and the Danish ones (1) very well, (2) satisfactorily, (3) not well at all. This inconsistency brought all Finnish responses to the first two mentioned categories.

The hypothesis that research workers in large cities with a relatively easy access to sources of information and more abundant contacts with colleagues would estimate their ability to keep up with advances to be better than those working in the country side and in smaller cities, was not supported. No difference was found between the responses from these groups. It should be pointed out, however, that the majority (90%) in the Danish group work in the Copenhagen area, while only 60% of the Finnish respondents work in the capital Helsinki.

III. ESTIMATE THE RELATIVE IMPORTANCE TO YOU OF THE FOLLOWING SOURCES OF INFORMATION: (1) LITERATURE AND (2) VERBAL SOURCES (QUESTION 4)

The estimated minimum percentage of information obtained from the literature was 20% and the maximum 100%. As illustrated by Table 4 the median percentage was 75% —considerably higher than that found by Herner (8) and Hertz and Rubenstein (9) among similar American populations. This seems to manifest either a greater literature-mindedness or a lower rate of direct communications among Scandinavian research workers than among their American colleagues. Most probably both factors play a role. The upper quartile was 90% in all sections save the Finnish academic one, in which it was 100%, and the lower quartile was about 60% in all sections save the Danish research section and the Finnish academic section, which both had a lower quartile of 70–75%.

TABLE 4. *Relationship between estimated percentage[a] of information obtained from the literature (as opposed to verbal sources) and the type of institution (Question 4; based on 187 responses)*

	Percent												
Institution	*100*	*90*	*80*	*75*	*70*	*60*	*50*	*40*	*30*	*25*	*20*	*10*	*Total*
Danish													
Academic, no.		7	7	1	3	1	3	2		2			26
Cumul., %		27	54	58	69	73	85	92		100			
Research institutes, no.	2	6	3	9	2	1	5	1					29
Cumul., %	7	28	38	69	76	79	97	100					
Industrial, no.	2	7	4	5	4	2	9		4	1			38
Cumul., %	5	24	34	47	58	63	87		97	100			
Total Danish cumul., %	4	26	41	57	67	71	89	92	97	100			93
Finnish													
Academic, no.	5	4	3	2	1	1		2		1			19
Cumul., %	26	47	63	73	79	84		95		100			
Research institutes, no.		7	3		7	5	2		4				28
Cumul., %		25	36		61	79	86		100				
Industrial, no.	4	12	5	7	1	3	8	4	2		1		47
Cumul., %	9	34	45	60	62	68	85	94	98		100		
Total Finnish Cumul., %	10	34	46	55	65	75	85	92	98	99	100		94
Both groups Cumul., %	7	30	44	56	66	73	87	92	98	99	100		187

[a] No correlation was found between the estimated percentage of information obtained from the literature and the location of institution (large city or rural).

It is interesting to note that in the similarity of the answers by the two groups from different countries the only significant difference appears in the heavy (90–100%) dependence on literature manifested by a greater part of the Finnish group than of the Danish one. In both academic sections the high dependence on literature was more common than it was in the other sections, as Herner (8) has discovered. There was little difference in this respect between the industrial and research sections. The low degree of dependence on literature—40% or less—was reported by 3–16% of the respondents scattered rather evenly over both groups and all sections.

Table 5 shows the answers to the same question tabulated according to the field of research. The greatest use of literature was made by pure mathematicians and physicists, pure chemists, and pharmaceutical scientists, and there was a very small, statistically insignificant, difference between these groups and the other subject groups in contrast to the findings of Herner (8). In all subject groups save the earth scientists and biologists the dependence on literature as opposed to verbal sources was greater than that found in the American study (8). The 17 workers in the fields of wood, rubber, paper, textile, and plastics technology reported the lowest degree of dependence upon literature, and remarks to the effect that research results were kept secret and seldom promptly published accompanied several of these responses.

Apart from the last-mentioned group little difference was found in the percentages of information obtained from the literature by pure scientists on one hand and the applied scientists on the other. In view of Herner's findings, this similarity again seems to indicate that both groups in the present study, especially the Finnish ones, perhaps are less communicative in their research work than their American colleagues.

TABLE 5. *Relationship between estimated percentage of information obtained from the literature and the field of research (Questions 1 and 4; based on 182 responses)*

Minimum % obtained from literature	*Cumulative %*								
	51/53[a]	*54*	*55/59*	*61*	*62*	*63/65*	*66/664*	*665/669*	*674/678*
25–40	100	100	100	100	100	100	100	100	100
50–60	100	82	75	80	95	64	88	97	82
70–80	86	76	50	80	70	45	71	66	53
90–100	36	35	25	40	30	45	29	34	12
Total No.	14	17	12	5	40	11	34	32	17

[a] UDC classes specified under Table 1.

IV. ESTIMATE THE TIME YOU SPEND WEEKLY ON SEARCHING AND READING LITERATURE (QUESTION 5)

Table 6 illustrates the dispersion of estimates which ranged from 1 to 50 hours per week spent on literature. In the Danish group there were more respondents limiting their reading to 1–2 hours/week and less respondents devoting over 12 hours to it, and the Danish average time 8.9 hours was thus slightly shorter than the Finnish 9.7. The Danish median was between 7 and 8 hours and the Finnish median between 9 and 10.

All these figures are higher than those reported by Shaw (17) (2 hours/week on the basis of the diaries and 4–5 hours on the basis of the answers to the questionnaire), by Bernal (1) (5.3 hours and only 4.9 hours by junior scientists), by Thorne (20) (5.1 hours on reading+0.6 hours on searching), and by the Cleveland study (14) (4.5 hours).

There are probably several reasons for this difference in estimates, among which the following factors regarding the Scandinavian groups should be pointed out: *(a)* the relatively high dependence on literature as a source of information as opposed to verbal sources, *(b)* by far the greatest part of literature is in some foreign language and is likely to require somewhat more reading time; *(c)* the scarcity of literature scientists or documentalists, who would make a selection to weed out unnecessary material, and the low degree of exploitation of such existing services. There was a significant difference between the Danish and Finnish industrial sections. The average time spent by the former was 7 hours/week or 2.5 hours less than the latter. What causes this difference cannot be judged from the available data.

TABLE 6. *Relationship between estimated time spent on literature and institutional affiliation (Question No. 5; based on 181 responses)*

Institution	*1–2*	*3–4*	*5–6*	*7–8*	*9–10*	*11–12*	*13–14*	*15–16*	*17–18*	*19–20*	*Over 20*	*Average*
Danish (89)												
Academic	3	2	3	7	5	2		3				8
Research institutes	1	1	5	4	5	4		4		3	1	11
Industry	3	4	11	6	8	2		1		1		7
Total Danish	7	7	19	17	18	8		8		4	1	8.9
Finnish (92)												
Academic	1	4	2	1	4			4		2	1	10.5
Research institutes		6	5	3	4	3	1	3		1	1	9.5
Industry	2	3	11	6	8	5	2	5		3	1	9.5
Total Finnish	3	13	18	10	16	8	3	12		6	3	9.7
Total	10	20	37	27	34	16	3	20		10	4	9.3
	37.0%			42.6%			20.4%					

Table 7 shows a direct correlation between the estimations of time devoted to the literature and the percentage of information obtained through reading.

V. HOW MANY SCIENTIFIC AND TECHNICAL JOURNALS AND SERIES DO YOU PERSONALLY SUBSCRIBE TO AND OBTAIN THROUGH MEMBERSHIPS? (QUESTION 6a-b)

As shown by Table 8 the average number of journals received personally by the respondents was 3.1. This figure checks reasonably well with the finding of Bernal (1) (2 journals, only 1.1 by junior scientists), of Herner (8) (2 journals), of Törnudd (21) (4.1), and of Shaw (17) (2.1). Institutional affiliation did not appear to affect personal subscriptions.

TABLE 7. Relationship between estimated percentage of information obtained from the literature and the estimated time devoted to it (Questions 4 and 5; based on 180 responses)

Minimum time hours/week	*Cumulative %*				*Number*
25–40	*50–60*	*70–80*	*90–100*		
1–3	100	100	100	100	18
4–6	83	93	95	85	49
7–10	54	67	69	57	60
11–14	8	18	42	32	19
15–18	8	5	27	24	21
19–	0	0	8	15	13
Total No.	24	39	64	53	180

TABLE 8. Number of journals personally subscribed to or received through membership in scientific societies (Question 6a-b; based on 84 Danish and 85 Finnish responses)

Institutional affiliation	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*Over*	*Average*
Danish (84)										
Academic	3	2	7	5	2	1	2	1		2.7
Research institutes	1	4	6	2	3	6	1	1	3	3.8
Industrial	1	5	8	8	2	4	2		4	4.1
Finnish (85)										
Academic	1	6	2	1	5	2	1			2.7
Research institutes	1	4	8	5	4	3	1			2.7
Industrial	3	8	12	9	3	2	3		2	2.7
Total	10	29	43	30	19	18	10	4	7	3.1

VI. HOW MANY SCIENTIFIC AND TECHNICAL JOURNALS AND SERIES DO YOU FOLLOW REGULARLY? (QUESTION 6c)

Table 9 illustrates the dispersion of the number of journals read or scanned ranging from 1 to 50. The average number is as high as 18 and median about 15. In accordance with findings mentioned under Sections III and IV the number is considerably higher than that found in previous studies. Bernal reported 9.7 journals for the whole sample and 7.1 journals for the group of junior scientists and Törnudd 13.2 journals.

Among British technologists employed by the electrical industry some 10% of the sample said they saw no journals regularly, about one-third claimed to see between 1 to 3, a further one-third between 4 to 6, 17% between 7 and 11, and 6% more than 12 (27a). These figures are, however, not comparable with the previously cited ones, because they refer to industrial technologists of various kinds the majority of whom did not have academic degrees.

TABLE 9. *Number of journals regularly read (or scanned) (Question 6; based on 173 responses)*

Institution	*1–5*	*6–10*	*11–15*	*16–20*	*21–25*	*26–30*	*31–35*	*36–*	*Total*	*Average*
Danish										
Academic	2	6	5	1	1	5	1	3	26	20
Research institutes	3	6	4	3	2	3	2	1	25	16
Industrial	4	9	8	9	1	2		3	36	17
Total Danish	9	21	17	13	4	10	3	7	87	18
Finnish										
Academic	3	5	5	1	1		1	1	17	14
Research institutes	2	8	1	5	5	1	3		25	17
Industrial	3	6	12	14	3	3		4	44	18
Total Finnish	8	19	18	20	9	4	4	5	86	17
Both groups									173	17
Cumulative, %	100	89	66	46	27	19	11	7		

VII. WHICH OF THE FOLLOWING TYPES OF PUBLICATIONS DO YOU USE (A) FREQUENTLY, (B) OCCASIONALLY, (c) SCARCELY EVER? (QUESTION 7)

Table 10 gives percentages of respondents using the six given types of publications and in parentheses the percentage using them frequently. The order of importance of these publications was: journals, books, abstracts and indexes, reviews, research reports as individual publications, and patents in both national groups and in all sections save the Danish industrial one in which patents and research reports changed places.

The fact that there was little difference in the relative importance laid on the various publications among workers in different organizations and that monographs, handbooks and compendia were rated as nearly equally useful as journals is most interesting. The latter finding, also emphasized by Herner (8), Moss (27a), Thorne (20), and Törnudd (21), suggests that more attention should be paid to the characteristics and effectiveness of the nonperiodical literature and its documentation to satisfy the requirements of scientists.

About 20% of the respondents reported that they scarcely ever used abstracts or indexes. This figure checks very well with the finding by Glass (6) who found that more than 10% of the biologists belong to the non-users of abstracts, and with the University of Michigan study of physiologists (22) which revealed that 13% of the scientists referred to abstracts to little or no extent. Bernal (1) reported that 87% of the British sample of scientists used abstracts and the survey of mathematicians (19) showed that at least two-thirds of the mathematicians made some use of abstracts and that the use among those with a high publishing record was more frequent. There is no question of the fact that the overwhelming proportion of research workers in all scientific fields make some use of abstracts, while their use by, for example, industrial technologists is smaller. Moss (27a) reported that only about 25% of the technologists in the British sample of technologists employed by electrical firms had made any use of abstracts in the 3 months preceding the enquiry and that 63% did not know about abstracts.

TABLE 10. Relative importance of different publications according to institutional affiliation (Question 7; based on 188 responses) (Figures in parentheses give % using the publication frequently)

	Percent Using					
Institution	*Journals*	*Abstracts & indexes*	*Reviews*	*Books*	*Patents & specifications*	*Research reports*
Danish						
Academic	96	85	56	85	22	48
	(89)	(48)	(30)	(70)		(15)
Research institutes	97	72	72	93	55	69
	(86)	(45)	(21)	(62)	(7)	(17)
Industrial	100	79	68	97	58	53
	(89)	(34)	(5)	(61)	(21)	(16)
Total Danish	98	79	66	93	47	56
Finnish						
Academic	100	89	69	100	26	63
	(89)	(63)	(11)	(74)		(16)
Research institutes	100	85	64	100	46	68
	(83)	(46)	(7)	(68)	(14)	(14)
Industrial	100	87	56	100	44	66
	(94)	(34)	(13)	(64)	(4)	(6)
Total Finnish	100	87	61	100	41	66
Both groups	99	83	63	97	44	61

Urquhart (23) found that abstracts were the main sources of citations used by the borrowers of the Science Museum Library. This finding reflects the residual nature of the material borrowed from this central library. The finding by Williams (26) that references in journal articles and books were used more frequently than abstract publications in the John Crerar Library cannot be compared with Urquhart's finding, because it excluded references to recent material which made 50% of the citations in Urquhart's sample. Bernal (1), on the other hand, found that footnotes were used as sources of reference more than twice as often as abstracts by scientists who did not have an opportunity to make use of an efficient information service.

The fact reported by Shaw (17) and Thorne (20) that only 5–6% of the reading acts concerned abstracts indicate that abstracts do not belong to the favorite pastime readings of scientists. The findings do not, however, mean that the ratio of their value to the value of, say, periodicals would correspond to the respective percentages of reading acts found by analyses of diary studies.

Shaw (17) reported that 14–18% of the literature read by scientists at the Forest Products Research Laboratory was report literature and Thorne (20) that the corresponding percentage was 44%. Herner (8) and Törnudd (21) also found a considerable dependence on reports. The present result, indicating a minor use of research reports, mirrors the fact that reports as a publishing media are less frequently used in the Scandinavian countries. Foreign reports, especially classified ones, scarcely enter the scene, and form a still greater problem in small countries than in the USA and the UK. Joint efforts in the documentation of report literature are urgently called for to supplement programs like the documents exchange scheme operated by the European Productivity Agency.

The patents, standards, and specifications formed the least important class of publications in the present study, and there is reason to assume that especially patents are frequently overlooked in the course of research. This seems to be due to unfamiliarity with patent indexes and the free services offered by national patent offices. The liaison between the last mentioned and the industry has received special attention in Denmark and might account for the slightly greater use of patents in the Danish group. In the EPA study (18) 12% of the US firms, 10% of the German ones, 7% of the Italian and British ones, and 8% of the Norwegian ones had used patent specifications to solve major problems. Patents were, however, lowest on the list of different types of publications save commercial reports.

Two-thirds of the research workers in the present study had made at least some use of reviews though only about one-tenth had used them frequently. These figures check well with the Bernal's (1) corresponding figure, 76%, and with the figure 77% obtained in the study of American physiologists (22). The survey of mathematicians, however, revealed that reviews were as heavily read as research papers, that is by *ca.* 85% of the total number of respondents. The high quality of *Mathematical Reviews* seems to account for this fact.

VIII. LIST THE FOREIGN LANGUAGES THAT YOU CAN READ (QUESTION 8)

Table 11 shows that all the Scandinavian respondents are able to read technical literature in Danish, Norwegian, Swedish, German, and English. Three-fourths of the respondents in the Danish group read French. The corresponding figure in the Finnish group is only one-fourth because Finnish constitutes an additional language either as mother tongue or second domestic language. This language record is higher than any previously reported corresponding data, simply because research workers in small countries with less known languages have not been involved in earlier studies. The weaknesses of the Scandinavian group are to be found in the eastern languages, Russian, and Spanish.

Bernal (1) found that 47% of the British research workers did not easily read any foreign language, and Shaw (17) reported that only 2.3–4.8% of all publications read were in another language than English. The survey of mathematicians (19) revealed the astonishing fact that one-seventh of the respondents were able to read Russian and that only about 4% had encountered difficulties because of inability to read another foreign language.

TABLE 11. Number of foreign languages[a] read (Question 8; based on 185 responses)

Nationality	*2*	*3*	*4*	*5*	*6*	*Over 6*
	Cumulative %					
Danish (92)	100	73	23	3	1	
Finnish (93)	100	98	26	6	3	1

Percentage of respondents reading different foreign languages

	Danish %	*Finnish %*
Danish		100
Finnish	2	(100 % of the Swedish-speaking respondents in the group)
Norwegian	100	100
Swedish	100	100
German	100	100
English	100	98
French	74	23
Dutch	12	5
Italian	12	5
Spanish	2	5
Russian	3	1
Portuguese	0	1
Esperanto	0	1

[a] Swedish, Norwegian, and Danish were counted as one language. These Scandinavian languages were considered foreign only to the Finnish respondents with Finnish as their mother tongue. Finnish was counted as foreign to all but the last-mentioned category.

The following figures reported by the EPA study (18) as percentages of small and medium sized firms in which foreign language articles had been read in the last year illustrate the position of small countries with little known languages: Austria 30%, Belgium 72%, Germany 30%, Norway 77%, UK 14%, and USA 9%.

In view of the increasing importance of the Russian, Japanese, and other eastern languages a rational solution to the problem seems to be close cooperation between the large language groups which might take the form of exchange of translations or at least abstracts in each others' languages. The often heard plea that scientists should learn more languages cannot but with difficulty be applied to small countries where the number of necessary languages already at present is high, and where contributions to international cooperation are currently made in that original research is published in a world language rather than in the national languages.

IX. DO YOU KEEP A PERSONAL INDEX? (QUESTION 9)

As seen from Table 12 more than 50% of all respondents and 60% in the Danish group maintain personal indexes. These figures are in agreement with 65% found by Bernal (1) and 64% found by Törnudd (21).

TABLE 12. Keeping personal indexes (Question 9; based on 188 responses)

	Danish group (94)		*Finnish group (94)*		*Both groups*	
Institution	*Number*	*%*	*Number*	*%*	*Number*	*%*
Academic	19	70	10	53	29	63
Research institutes	18	60	15	54	32	57
Industrial	20	51	26	55	45	54
Total	57	60	51	54	106	57

X. DO YOU HAVE AN OPPORTUNITY TO USE REFERENCE AND INFORMATION SERVICES OF A LIBRARY? (QUESTION 10)

As seen from Table 13 about three-fourths of the research workers indicated that they had an opportunity to resort to the services of a library. That this opportunity is inefficiently exploited or that the services are not as adequate as one might expect will be illustrated under the next heading.

No difference was found between the responses from scientists working in large cities and those working at a distance from these centers.

TABLE 13. Opportunity to use reference and information services of a library (Question 10; based on 188 responses)

Institution	*Number*	*%*
Danish		
Academic	19	70
Research institutes	26	90
Industrial	32	84
Total Danish	77	82
Finnish		
Academic	11	58
Research institutes	19	68
Industrial	39	83
Total Finnish	69	73
Both groups	146	78

Moss (27a) found that rather less than a tenth of the technologists employed in British electrical firms with 200–1000 employees appeared to work in an organization where there was a specific, named official with the function to draw the attention of scientific and technical staff to useful information, and the EPA study (18) revealed that the majority of small and medium sized establishments had not consulted any kind of library during the last year, the percentages varying from 63% in the UK to 92% in Italy.

XI. WHICH OF THE SERVICES LISTED BELOW HAVE YOU MADE USE OF: (A) FREQUENTLY, (B) SOMETIMES, (C) NEVER? (QUESTION 11)

(a) Loan and photocopy services constitute the only form of service which has been used by the majority of the scientists in both national groups. That the given percentages are not 100 does of course not imply that the remaining percentage would not have used a library. It merely implies that the libraries used have not offered personal services to the reader.

(b) Quick reference service is frequently used by less than a sixth of the re-

TABLE 14. *Use of different library services (Question 11; based on 186 responses)*

Institution	Percentage using each form of service[a]										
	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*i*	*k*
Danish (92)											
Academic	95 (88)	37 (15)	19 (4)	7	4	15 (4)	11 (4)	4	29 (7)	33 (11)	67 (11)
Research institutes	96 (75)	57 (11)	43 (4)	18	11	14	29	25 (7)	33 (4)	43 (7)	79 (25)
Industrial	100 (65)	49 (14)	40 (5)	22 (3)	14 (3)	24 (5)	14	5	22	14	70 (5)
Total Danish	98 (75)	48 (13)	35 (4)	16 (1)	10 (1)	18 (3)	17 (1)	11 (2)	27 (3)	28 (5)	72 (13)
Finnish (94)											
Academic	94 (47)	21	10 (5)			5		5	32 (16)	11	37 (16)
Research institutes	75 (64)	57	14	4		15 (4)	39	7	53 (14)	36 (11)	25 (7)
Industrial	92 (60)	47 (4)	17 (4)	4		4	34	9	19 (4)	18 (9)	21 (2)
Total Finnish	87 (59)	45 (2)	15 (3)	3		7 (1)	29	7	32 (10)	21 (6)	26 (6)
Both groups	92 (67)	46 (8)	25 (4)	10 (1)	5 (1)	13 (2)	23 (1)	9 (1)	30 (6)	25 (6)	50 (10)

[a] Percentage frequently using the form of service in parentheses.

a, loan and photocopy services; *b*, quick reference service; *c*, brief literature searches; *d*, comprehensive literature searches; *e*, critical searches with an evaluation of the information reviewed; *f*, continuous scanning of the literature and "feeding" relevant items; *g*, translations of foreign language publications; *h*, abstracting papers specified by the client; *i*, translation of the client's publication into a world language and checking the linguistic form; *j*, editorial assistance; *k*, guidance by the library staff.

spondents and has been used sometimes by about 50% of the respondents if the academic sections, who had availed themselves of this service to a still smaller extent, are omitted. This finding must be considered shocking in view of the time lost by research workers, who apparently have to leave their work and go to a library to check every simple bit of information. Unfortunately no comparative data from other studies are available. The author's limited experience with reference services in American and British libraries, both public, university, and special libraries, however, suggests that the facilities and their utilization in these countries are of another magnitude than they are in Scandinavia, where the majority of libraries still concentrate their activities on the classic aspects of librarianship. This does not mean that there would not be efficient information centers in Scandinavia, but the results prove that the services offered by these centers are not utilized by younger research workers. Whether this deficiency is due to the inability of the centers to “sell” and publicize their services or to a lack in the education of research workers or their personal preferences, is difficult to decide. Probably all aspects need attention and urgent measures in view of the shortage of qualified researchers.

(c) Brief literature searches on request have been made for a still smaller proportion of the respondents, for 35% of the Danish and for 15% of the Finnish group. Here again the academic sections have sought the service to a smaller extent than the others in accordance with the findings of Herner (8). This difference is naturally due to the fact that brief literature searches are seldom needed by a person working on long-term projects of specialized nature. The above-mentioned percentages can be compared with Herner's 65% referring to the total sample at Johns Hopkins and 35% referring to the pure scientists versus 79% to the applied scientists. It should be noted, however, that the last-mentioned figures cover both brief and more comprehensive literature searches.

The conclusions made under (b) apply to these services and the laments can rightly be raised to the second power.

(d) Comprehensive literature searches on request belong to less self-evidently needed services, because it can be claimed that a research worker gains substantially more by performing a large part of the search himself. This applies especially to exhaustive searches with the aim of reviewing a field to be entered in the course of a new research project. Patent searches concerning patentability, state-of-the-art, infringement, and validity, on the other hand, undoubtedly belong to the domain of information specialists, as well as, for example, bibliographies requested by research workers who are preparing material for publication.

The results of the present survey show that a very small proportion of the respondents have had occasion to have comprehensive searches made: 16% of the Danish group and only 3% of the Finnish one. Herner's study on the service requirements among research workers at Atlantic Research Corporation (8a) revealed that comprehensive literature searches were in high demand, and there is no reason to assume that the Scandinavian respondents would not profit by having more searches made.

(e) Critical surveys of the literature searched on request in which attention is paid to the value of reported results had been carried out for 10% of the Danish group, mainly for the industrial employees, while nobody in the Finnish group had used this type of service which requires a very highly qualified searcher.

It need not be considered Utopian to make a plea for improved facilities for this type of service also, although its development demands an effective program for the training of documentalists or information specialists.

(f) Continuous scanning of the literature on a subject requested by the research worker and the "feeding" of references, abstracts, or the material as they appear is one of the most important tasks of any information service, and the closer the contact between the service point and the research worker the more efficiently this service can be rendered. Therefore it is deplorable to note that only 18% of the Danish research workers and 7% of the Finnish ones had ever made use of this service. It is out of the question, of course, that the research worker by this means could be supplied with all the information he needs, but there is no doubt that a research worker's burden could easily be lightened and that he or she would escape scanning some of the worthless material accumulating if there were a selective feeding service, designed to meet the requirements.

(g) Translations of foreign language publications into mother tongue were obtained by 17% in the Danish sample and by 29% in the Finnish group. These figures are lower than 33% found by Herner (8) and about 50% found by Törnudd (21) in studies of American scientists. The translations reported were mainly from Russian and in the Finnish group also from French.

(h) Abstracting articles specified by the research worker had been carried out for 11% in the Danish group and 7% in the Finnish one.

(i) Translations of manuscripts into a world language and checking the language before publication were services mostly used by the academic and research sections in both national groups. On an average 30% of the respondents had had occasion to use this service and especially the checking service. This percentage covers a great majority of those who reported having published research results in a world language. There is every reason to improve these facilities, because the value of the additional effort of publishing original research in a world language is decreased by incorrect translations and linguistic oddities.

(j) Editorial assistance in proofreading, diagrams, etc., had been received by about one-fourth of the respondents or by about one-half of those who had made contributions to the literature.

(k) Guidance by the library staff when searching the literature had been resorted to by almost three-fourths in the Danish group but only by one-fourth in the Finnish one. The Danish figure is in agreement with the corresponding 70% reported by Herner (8), while the Finnish figure is surprisingly low. The questions were framed in similar ways in both languages and cannot have caused the difference.

XII. IF YOU HAVE NOT HAD AN OPPORTUNITY TO AVAIL YOURSELF OF SOME SERVICE LISTED ABOVE WHICH YOU FEEL A GREAT NEED FOR, MARK IT WITH AN ASTERISK (NOTE: ONLY ONE ITEM) (QUESTION 12)

As seen from Table 15 only 66 respondents indicated a requirement for some additional service. Some respondents checked several items, although it was pointed out that only one check was wanted, and in these cases all but the simplest type of service checked were omitted.

Continuous scanning of the literature and feeding of pertinent material received the greatest number of votes and comprehensive literature searches were second. The other services that were checked as needed were in this order: brief literature searches, translation of manuscripts into world languages, translation of foreign language publications, critical literature surveys, and abstracting specified articles. Quick reference service was missed by only 2 respondents,

TABLE 15. *Unavailable services for which the need is greatest (Question 12; based on 66 responses)*

	Number of respondents requiring the service[a]										
Institution	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*
Danish (29)											
Academic				2		4	2	2	3		
Research institutes		1		1	1	1	1				
Industrial			1	4	1	3		1	1		
Total Danish		1	1	7	2	8	3	3	4		
Finnish (37)											
Academic			1	2		2	1		2		
Research institutes	2		3	3		6					
Industrial		1	1	4	2	6	1				
Total Finnish	2	1	5	9	2	14	2		2		
Both groups of respondents, %	3	3	9	24	6	33	8	5	9	—	—

[a] The services a–k are specified under Table 14.

and two Finnish respondents expressed their need for loan and photocopy service.

It seems that the comparatively large amount of time spent by the respondents and the relatively large number of journals reported to be regularly read or scanned in this study and the previous study by Törnudd (21) are due partly to the fact that library services have not been available or have not been utilized efficiently. This interdependence was also suggested by Bernal (1). Assuming that three hours, a little more than one-half of the difference between the average time spent by the respondents in this study and that reported in the other British and American studies, could be saved for the research worker, it would be economically profitable to employ one literature specialist for every 9 research workers, who still would spend 6–8 hours on the literature.

XIII. HOW MANY PAPERS, RESEARCH REPORTS, AND BOOKS HAVE YOU PUBLISHED DURING THE PAST THREE YEARS AND IN WHICH JOURNALS HAVE YOUR PAPERS APPEARED? (QUESTION 13)

As seen from Table 16 the average number of papers (books counted as three papers) amounted to 3.6 and the average number of papers among those who had published something 5.8. The corresponding medians are 1 and 4. The publishing record in the industrial section is lower than that in the academic and research sections. It must be pointed out that only about one-half of the publications were actual research papers, and the rest articles in trade journals.

The results agree with those obtained in the American study of physiologists (22) who were reported to publish on an average 4–5 papers in three years and among whom young physiologists up to 39 years of age mostly published a maximum of four papers. The survey of American mathematicians (19) revealed

TABLE 16. *Number of papers published during the past 3 years according to institutional affiliation (Question 13; based on 188 responses)*

Institution	*No answer*	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11–15*	*16–*	*Total average*	*Average by author*
Danish																
Academic		3	1	4	3	2	2	2	2	2		3	1	2	6.6	7.5
Research institutes		9	2	3	4	2	4	1	1	1			1	1	4.5	6.6
Industrial	3	25	2	2	5			1						.	0.8	2.7
Total Danish	3	37	5	9	12	4	6	4	3	3		3	2	3	3.7	6.2
Finnish																
Academic	2	3	4	1	2	2	2				1		2		3.9	4.7
Research institutes	1	5	3	2	3	5	5	1					1	2	4.8	6.0
Industrial	12	20	6		2	2		3					1	1	2.1	5.0
Total Finnish	15	28	13	3	7	9	7	4			1		4	3	3.4	5.3
Both groups	18	65	18	12	19	13	13	8	3	3	1	3	6	6	3.6	≥5.8

that 85% of the sample produced one publication or less a year and accordingly a maximum of 3 publications in three years. The top 15% produced 1.1–2.3 publications a year or 3.3–6.9 publications in three years. Scates and Yeomans (16) report that 139 employees at New York Naval Shipyard published on average 0.1 article during three years, but this low figure is due to the fact that only 9 persons answered the question. Among the last-mentioned respondents the average number of publications was 2.1.

As seen from Table 17 three-fourths of the Finnish publications for which media were specified appeared in domestic journals, while more than one-fourth of the Danish contributions were published in foreign or international journals and another fourth in joint Scandinavian periodicals. This implies that at least 50% of the Danish publications appeared in world languages. The percentage of Finnish contributions published in world languages was certainly higher than the 25% published in non-domestic periodicals because several Finnish journals containing both Finnish and foreign language articles were specified.

TABLE 17. Nature of publishing media used for own papers (Question 13; based on 358 publications for which media was specified by authors)

Nature of journal	*Danish group, no. of papers*	*Finnish group, no. of papers*	*Both, % of papers*
National	80	155	66
Joint Scandinavian	37	23	17
Foreign	45	18	18
Total	162	196	

XIV. WHAT ARE SOME OF THE DIFFICULTIES YOU HAVE HAD WITH YOUR OWN PUBLICATIONS? (QUESTION 14)

Only 42 respondents specified problems they had encountered in connection with their own publications as seen from Table 18.

Slowness in publication, one of the suggested problems, was written in by a good third of the respondents, while lack of suitable journals was mentioned by only one-sixth, namely by research workers in very special fields. Lack of financial support for printing and translating manuscripts into a world language appears to have hampered the publishing activities of a large enough group to suggest a study of the circumstances. Lack of time was mentioned by one-fourth of the respondents and lack of clerical and editorial help by one-sixth.

Summing up, it seems that the Scandinavian group of junior research workers has reasonably adequate opportunities to get their contributions published. The fact that a large part of original papers is published in internationally lesser known domestic periodicals seems to constitute a more difficult problem, as the contributions often remain unnoticed by the international abstracting and indexing publications. This fact was recently revealed in a Norwegian study carried out by the Interim Abstracts Committee of the Joint Committee of the Norwegian Research Councils.

TABLE 18. *Difficulties with own publications (Question 14; based on 24 Danish and 18 Finnish responses)*

Difficulties mentioned (first four problems given in the questionnaire)	*Danish group*	*Finnish group*	*Both, % of respondents*
Lack of suitable journals	4	3	17
Slowness in publication	11	4	36
Lack of financial support for printing	6	6	29
Lack of financial support for translation into a world language	5	7	29
Lack of time	4	7	26
Lack of competent translators and language examiners	5		12
Lack of clerical and editorial assistance	3	3	14
Lack of know-how in reporting	2		5
Restrictions by the company		2	5
Other: necessity to popularize, time lag required for wetting, inability of printers to set mathematical text	2	1	7

That research workers employed by industry have special problems with their publications is evident from their low publishing record reported above and revealed in the study of physiologists (22). Measures toward decreasing unnecessary secrecy are called for.

XV. WHAT ARE SOME OF THE DIFFICULTIES YOU HAVE IN OBTAINING REQUIRED INFORMATION AND KEEPING UP WITH ADVANCES (QUESTION 15)

As seen from Table 19 lack of time constitutes the most common problem (71%) followed by lack of access to published material (23%), lack of suitable abstract journals and reviews (16%), lack of adequate library services (14%), and slowness in publication (6%). This list of problems gives quite another picture than that obtained in the American study of physiologists (22). The main problems among the latter population were lack of time (48%), too many publications and too much content in the field (30%), lack of access to published material and inadequate library facilities (11%), slowness in publication (8%), and inadequacy of the content of published material (7%).

The difference seems to be due to less adequate library facilities at the disposal of the Scandinavian groups.

Slowness in publication indicated as a major problem by relatively few respondents in these studies was considered a more serious obstacle by mathematicians in the American study (19) which revealed that one-sixth of the respondents had needlessly duplicated the work of others or otherwise been hampered in their research because of long delays in publication of the work of others. 25% of the respondents indicated that this problem was a really serious one.

TABLE 19. *Difficulties connected with acquiring needed information and keeping up with advances (Question 15; based on 76 Danish and 72 Finnish responses)*

Difficulties mentioned (first four problems given in the questionnaire)	*Danish group*	*Finnish group*	*Both, % of respondents*
Time limitations	64	53	71
Lack of access to published material	16	18	23
Slowness of publications	8	1	6
Lack of suitable abstract journals and reviews	13	11	16
Time lag in abstracting	2	3	3
Too many duplicating abstract journals	2		1
Inadequate documentation of theses	2		1
Lack of library services and loans	6	6	8
Lack of experience of and belief in library services	1	1	1
Slow deliveries of loans & photocopies from central libraries and from abroad	1	7	5
Too large field and too much published, rehashes	2	3	3
Inadequacy of content of published material, untrustworthiness	2	1	2
Inadequacy of presentation of material: frequent use of block diagrams	1		1
Isolation from colleagues at home and abroad	6		4

Long delays in publication no doubt increase the probability of undesirable duplication of research. Accelerating the publishing speed would, however, solve but a fraction of the problem complex involving such additional factors as indicated above and exemplified under the following heading.

XVI. CAN YOU DESCRIBE SPECIFIC INSTANCES OF UNDESIRABLE DUPLICATION IN YOUR WORK CAUSED BY THE LACK OF INFORMATION ON RESEARCH CARRIED OUT ELSEWHERE? (QUESTION 16)

Only 19 cases listed in Table 20 were written in, and the small number does not encourage generalisations. It should, however, be noted that 2 of the cases were caused by delays in publication, 3 cases of duplication were due to industrial research results not published; at least 2 cases were caused by the fact that the previous results had been published in a relatively obscure publication. One of the cases was due to inefficient cooperation between similar institutions in neighbor countries. One of the cases was definitely caused by inadequate searching of the literature and one by the inadequacy of the publication in question. Whether the remaining 9 cases were caused by inadequate searching of the literature, lack of access to relevant publications, by inadequate indexing, by the literature itself, or by other factors cannot be judged on the basis of the obtained information. A repeated search of the literature on each of the subjects would answer these questions in regard to the documentation aspects of the problem, while the actual circumstances and thus the decisive factors might be overlooked.

TABLE 20. Specific cases of undesirable duplication of research carried out because information was not readily available (Question 16; 19 cases)

Chemistry

1. Investigation on a new complex builder was published, and a letter received to inform that the same study had been carried out in the USA.
2. The composition of cork was investigated for some time, before results of an identical study carried out in Spain became known, as their author supplied reprints of his papers.

Physics

1. Study on the validity of Lambert's law in regard to a photometer was duplicated before finding out that it had been performed in Germany.
2. 80 hours wasted time in the calculation of an optical filter before accidentally finding out that the problem had been solved and that the filter in question was commercially available.

Biochemistry

1. 6 months wasted on an investigation before results of an identical study appeared in *Nature.*
2. Development of analytical methods and studies on the structure of polysaccharides.

Mechanical and Electrical Engineering

1. Development of a casting process without knowledge of the fact that the process in question was patented.
2. Design of a frequency multiplier for FM radiophony.

Food Technology

1. Studies on the drying of lump sugar.
2. One month wasted in meat investigations, before an English publication on

the same research topic appeared. The latter had been submitted and accepted for publication by the English journal 10 months before it finally appeared.
3. Preparation of diatomaceous earth for filtering.
Chemical Technology
1. Manufacturing process for p-aminophenol was studied during the war. Later a detailed description of a commercially used method was found in a book.
2. An apparatus for peat gasification was designed in Finland and eventually an OEEC publication was found (Finland is not an OEEC country) in which an almost identical apparatus with similar dimensions designed at the University of Louvain was described. At the time this information was not available in any “ordinary” publication.
Ceramic Technology
1. Duplication occurred in ceramic investigations. The second study, however, lead to improved practical results.
2. Development of a method for the dispersion of sulphur in ceramic batches was carried out and considered a rather intricate problem, until it was found out that a preparation for the very purpose was manufactured by Imperial Chemical Industries Ltd.
3. Work on drying ceramic ware by a certain method was duplicated. Similar results of an American study were later found in a journal article.
Pulp and Paper Technology
1. Journal article mentioned an applicable test method without giving a literature reference. The method was never found described in the literature.
Building Technology
Moisture in cast roofs was studied simultaneously and independently by two Scandinavian research institutes.

Table 21 presents the number of affirmative answers to the question of unnecessary research performed. As could be expected duplication was reported by a greater proportion of industrial research workers than of scientists with academic and research institute affiliations.

TABLE 21. Number of respondents stating that specific instances of undesirable duplication of research has occurred because information was not readily available (Question 16)

	Danish group		*Finnish group*		*Both groups, %*
Institution	*No.*	*Specific cases described*	*No.*	*Specific cases described*	
Academic	4	3	2	2	13
Research institutes	5	3	4	2	16
Industrial	15	6	7	3	26
Total	24	12	16	7	21

Several respondents commented on the unnecessarily prolonged secrecy precautions which lead to needless mass duplication of research effort.

XVII. CAN YOU RECOMMEND IMPROVEMENTS WHICH MIGHT MAKE THE LITERATURE AND REFERENCE SERVICES MEET YOUR NEEDS MORE ADEQUATELY? (QUESTION 17)

Fifty-four suggestions listed in Table 22 for the improvement of literature and reference services were put forward. The great majority of suggestions concerned the periodical literature and were centered around weaknesses which are receiving attention but might require more efficient measures.

TABLE 22. Number of respondents presenting concrete suggestions for the improvement of the literature and the library services (Question 17; based on 188 responses)

Institution	*Danish group*	*Finnish group*	*Both, %*
Academic	10	5	33
Research institutes	8	8	28
Industrial	10	13	27
Total	28	26	29

List of suggestions made for the improvement of the state of affairs

Number of times made indicated with figure in parentheses.

Journals

a. *Grouping and reducing the number of journals* (8) with concentration of original work into international research journals rather than small national journals which preferably should specialize in review articles. International cooperation called for.

b. *Stricter editorial policy* (5) to avoid rehashes.

c. *Abstracts of all articles* (11) preferably printed on separate pages to facilitate cutting out and pasting on catalog cards.

d. *UDC classification* of American and English publications as well as others (7).

e. *More footnotes.*

f. *Standard size and running pagination of journals* (2).

Abstract journals and indexes

a. *Grouping and reducing the number of abstract journals* (3) through international cooperation.

b. *Speeding up abstracting* (6), e.g., by collecting of abstracts, immediately after manuscript has been accepted for publication in a journal.

c. *Fuller abstracts* (1).

d. *Need of fuller coverage* (4) was expressed for (1) theses, (2) Russian and Japanese literature.

- *e.* *Need for additional abstract journals* (3) was expressed in the following fields: fishery, materials of construction.
- *f.* *Editions of abstract journal printed on one side of the page* (1).

Reviews

- *a.* *Need for more critical reviews of the literature* (4) e.g., industrial management, operations research, and the handbook literature.
- *b.* *Need for critical reviews of commercially available instruments* and their application (2).

Libraries

- *a.* *More subject specialists in library staff* (6).
- *b.* *Improvement of subject catalogs in libraries* (3).
- *c.* More liberal lending policy and faster photocopying services (2).
- *d.* Under question concerning difficulties in keeping abreast of new developments inadequate holdings of libraries were pointed out by 34 respondents.

Information services

- *a.* *Tailor-made abstract services* with international coverage supplying abstracts from a specified field (9).
- *b.* *"University microfilms" for Europe* (1).
- *c.* *Development of existing national technical information centers* (2).
- *d.* *Documentation of initiated research programs* (1).
- *e.* *Translations of Russian and Japanese publications* (3).

Instruction in the use of libraries and in subject literatures

- *a.* *Courses in the use of libraries and subject literature on the undergraduate level* (5).
- *b.* *Formal instruction in the use of libraries at the high school level* (1).

Miscellaneous

- *a.* *Selling information-mindedness to the industrial management* (1).
- *b.* Publication form for unfinished researches which cannot be published in a normal way.
- *c.* Arrangements facilitating the use of colleagues' personal indexes.

It is notable that one-fifth of the respondents made a plea for the inclusion of abstracts of articles in all periodicals either in the form of index cards or printed in a way facilitating clipping. In addition to abstracts, the classification of articles and monographs by the Universal Decimal Classification was recommended. UDC is the most widely used system in Europe, and the often heard wish that American documentation would follow suit was made by several respondents.

To decrease the unnecessarily large number of research journals, suggestions were made to publish original research reports in international periodicals limited to narrow fields, rather than in local journals with limited circulation, which were considered more suitable, for example, for reviews. This trend is demonstrated by the great number of *Ada…Scandinavica* series established recently through the cooperation between scientific societies and research councils in northern Europe. These joint efforts have solved many problems in the dissemination of research results.

Regarding abstract journals, the speeding up of abstracting and international coordination to decrease the number of parallel services was emphasized by several respondents. Means similar to those developed by ICSU Abstracting Board were recommended for this purpose.

Concerning reviews, the need for critical reviews for horizontal subjects to supplement reviews of special fields was mentioned.

The concrete suggestions concerning libraries and information services covered the need for more subject specialists to carry out reference services in libraries and the improvement of subject catalogs. The latter requirement is notable as it does not come from librarians, as usually is the case.

The need for an abstract service currently supplying the client with cards from a selected narrow subject field was expressed by 9 respondents. It was pointed out that such services should be comprehensive and international in scope and preferably carried out through international cooperation. The idea is far from utopian. On a small scale it has been realized a.o. in Sweden where the Association for Documentation carries out an abstract service along these lines. The coverage of foreign material is selective and far from inclusive, while the Swedish scientific and technical publications are covered as fully as possible.

XVIII. HAVE YOU AFTER GRADUATION FELT A NEED TO IMPROVE YOUR SKILL IN THE USE OF LITERATURE?

Table 23 shows the responses to the above question which in more than 50% of the cases was affirmative. The difference between the answers by the Danish and those by the Finnish group do not bear any significance. They are namely due to an error in the questionnaire. The Finnish version of the question was erroneously formulated to read: Have you after graduation been compelled to improve your skill…, and this framing naturally decreased the number of affirmative answers.

In an American study (21) the percentage of respondents giving an affirmative answer to the same question was definitely lower: 44%. Williams in his study of patrons of the John Crerar Library (26) found that 68.8% of the respondents had had training in the use of libraries and 36.6% had received training in the literature of their field. In Scandinavian countries these matters have received very little attention, and there is every reason to believe that formal instruction in literature and library know-how is needed.

It is interesting to note the six suggestions made for the inclusion of courses in subject literatures and the use of libraries in the curricula of universities and institutes of technology and even at the highschool level. These were made in connection with suggestions for general improvements of the literature and library services.

TABLE 23. The proportion of respondents having felt a need for improving their skill[a] in the use of literature after graduation (Question 18; based on 176 responses)

Institution	*Affirmative answer, %*	*Negative answer, %*	*Total*
Danish			
Academic	44	66	27
Research institutes	71	29	28
Industrial	67	33	33
Total Danish	61	39	88
Finnish			
Academic	58	42	19
Research institutes	55	45	22
Industrial	36	64	47
Total Finnish	45	55	88
Both groups	53	47	176

[a] No correlation was found between the need for improving skill and the school of graduation.

SUMMARY AND CONCLUSIONS

The 72 studies summarized above furnish valuable information on the manner in which scientific literature and library services are used. Several recent studies have remained unpublished and deficiencies in the documentation of documentation research have rendered it difficult to trace relevant investigations. This fact is the more deplorable as the results obtained so far are fragmentary and the coverage of the studies far from adequate, not to mention the fact that the methodology for obtaining truly objective data still remains to be developed.

To summarize some of the findings made in recent investigations it should first of all be pointed out that in spite of the great diversity of information gathering behavior from scientist to scientist, there is a similar pattern to be found when scientists are studied in groups. This pattern has been proved to be influenced a.o. by the following factors:

1. The accessibility of information and the extent to which information services are available and utilized.

Improving library facilities and services should therefore result in a twofold improvement: in addition to receiving better service the scientist himself would certainly find more efficient ways of utilizing information.

A comparative study of scientists working in similar institutions with similar working conditions and staffs, and in similar subject fields but with different library facilities suggests itself for the measurement of the role played by services. Optimum service conditions applicable in setting up efficient services might be found by means of such comparisons.

2. The kind of work.

It has been found that research workers, teachers in academic institutions, and information specialists are the heaviest users of literature as well as the most diligent producers. The majority in this group rely on literature to a greater extent than on all other sources of information taken together. The number of persons in this group has recently been studied in most countries in view of the need to decrease the shortage of highly qualified scientists and engineers. That each member of this important group of scientific workers spends on an average from 2 to 10 man-hours a week on the literature means in terms of salaries and the manpower shortage a great investment rendering even minor improvements significant if they result in cutting down the time spent on literature or decreasing the unnecessary duplication of research.

Scientists employed by industry for other duties than research and development have been found to resort to the literature to a very small extent. Whether this fact is due to inadequate services and publications or to other factors requires clarification.

That available services are insufficiently utilized by most scientists regardless of their activities constitutes a grave problem requiring special attention. The aggressive dissemination of information by various services may prove to be excessive and too inefficient a method of work. Services rendered on request are likely to offer a better pay-off of the investment of time and money, provided requests are received.

3. The working environment.

The few studies which have attempted to throw light on the influence of institutional affiliation have concerned research workers and revealed that the greatest difference is probably to be found between scientists working in academic institutions on the one hand and those employed by industrial, government, and other research establishments, on the other. Whether the preference for performing all literature work personally, common in the first-mentioned group, is due to the environment or to the lack of adequate service facilities cannot be determined from the available data. That the latter group seems to be encountering greater difficulties should be taken into consideration.

4. The educational background.

A direct correlation has been found between the educational record and the extent to which literature is resorted to. This finding suggests the desirability of arranging courses in subject literatures and in the use of libraries for scientists with lower degrees. The fact that research workers and engineers have been found to lack skill in the use of different sources of information calls for the inclusion of such training in curricula of universities and institutes of technology in all countries. A recommendation to this effect from a suitable body should be considered.

5. The field of science.

Comparative studies have revealed that workers in the more exact sciences tend to depend on the literature to a still greater degree than, e.g., biologists and medical workers. It has also been found that the life of usefulness of publications varies from field to field. The finding that researchers in applied sciences would depend on the literature to a smaller degree than pure scientists disagrees with the results of the present Scandinavian study in which the scientists with only few exceptions used reference services. The excessive secrecy precautions concerning results of applied research do cause unnecessary duplication of research.

The finding that, e.g., American psychologists limit their information gathering mainly to material in their mother tongue while chemists and physicists use foreign publications more frequently seems to mirror the coverage of the respective abstract services as much as actual differences between the subject fields.

In short, the differences between requirements by workers in different subject fields may have been overemphasized.

6. Nationality.

In spite of the small number of studies carried out in other countries than the USA and the UK, it can be stated that nationality plays a certain role.

Language problems with respect to English, German, and French are of minor importance among research workers and other academically trained persons who are able to read these languages. An author's own publication that must be translated into a world language does, however, constitute an additional problem. Workers with a lower educational record are handicapped even with English, and require suitable publications in their mother tongue.

In small countries the relative importance of literature as opposed to verbal communications is naturally greater than in large countries, as the number of countrymen with the same speciality often is small, and foreign contacts are more difficult to make, especially by junior scientists. A more reserved general behavior pattern might also account for the decrease in importance of verbal communications as a source of information among non-American scientists.

There is no doubt that research workers in the USA and the UK enjoy information services seldom used by, although to some extent available to, Scandinavian scientists. Since the way in which small and medium sized industry uses technical and scientific information appears to be similar in the USA, UK, and European countries as revealed by the EPA study, the differences found between research workers from the various countries seem to be due to differences in the education of scientists.

Thus the exceptional features found in the Scandinavian group, such as the great amount of time devoted to literature, the large number of periodicals regularly read, the kinds of difficulties encountered in keeping abreast of new developments, and the small extent to which library services were demanded, may stem from the difference in emphasis laid on information gathering in connection with academic studies and the scarcity of information services, rather than on actual national factors.

7. The age of scientists.

It has been found that junior scientists use literature to a smaller extent than their seniors and that there is a curvilinear correlation between publishing activities and age, the most productive age being in the 40's. If the first-mentioned finding is typical it accentuates the exceptional features of the young Scandinavian group.

None of the methods used in studies of the use of information by scientists has proved to be truly reliable, and therefore the results of the operational research program underway in the USA are looked forward to with great expectations. Should these studies reveal better methods for the study of these problems, a coordinated research program carried out in several countries and covering a variety of users of information should be undertaken with the aim of collecting the relevant data needed to form a basis for the development of scientific literature and information services.

Meanwhile, local studies carried out by some of the methods directed to the users of information to clarify questions connected with services to smaller groups will be justifiable—in fact there does not seem to be any doubt of the necessity for every information service to carry out a continuing analysis of the requirements of its users, especially of its least industrious users.

ACKNOWLEDGMENTS

This study was made under a contract from the Department of Natural Sciences of the United Nations Educational, Scientific and Cultural Organization, and the fullest possible cooperation and help was given by Dr. J.E.Holmstrom.

Mrs. Helen L.Brownson, Program Chairman, Area 1 of the International Conference on Scientific Information, has shown kind interest in the work and sincere thanks are due for her invaluable help in collecting material. The authors of similar studies, completed and current ones, have been most helpful in supplying the author with both reports and advice.

The author wishes to express her thanks for the cooperation of the Junior Scientists Committee of the Danish Academy of Technical Sciences and their Finnish colleagues who were queried and who responded, sparing no trouble to bring the response rate up to nearly 100%.

Last, but not least, an acknowledgment is due to the Scandinavian Council for Applied Research for their interest in the project.

APPENDIX I. ENGLISH TRANSLATION OF THE CIRCULAR LETTER MAILED TO THE SCIENTISTS IN THE SAMPLE.

The use of scientific and technical literature and reference services

The rapid growth of the flow of literature has several harmful effects on technical progress: valuable research results often remain unexploited, research and development work is duplicated, and searching for as well as reading the professional literature requires more and more time of the busy research worker.

To help to improve the situation attempts have been made to find out how research workers obtain the necessary information and which are the greatest difficulties involved. In the USA several studies have been carried out with the aim of finding an answer to these problems, and Unesco has recently taken the initiative to perform comparative studies in Europe where these problems presumably are different.

Unesco has requested me to carry out a query in Scandinavia, and I have dispatched the enclosed questionnaire to a sample of 100 young research workers in Finland. With the kind permission of the Board of the Danish contact group Yngre Forskere I am now mailing the same form to a sample of 100 members of the group. I should very much appreciate it if you would be kind enough to answer as many of the questions as possible and return the filled in questionnaire. The study will be of little value if responses are not received from all.

The answers will be considered confidential and the results will be submitted in the form of statistics.

Sincerely yours

ELIN TÖRNUDD
Secretary of the Scandinavian
Council for Applied Research

Enclosure: questionnaire

APPFNDIX II. ENGLISH TRANSLATION OF THE QUESTIONNAIRE FOR RESEARCH WORKERS ON THE USE OF SCIENTIFIC AND TECHNICAL INFORMATION

You are requested to fill in this questionnaire and return it not later than 25/11/1957 in the enclosed envelope.

1. Present field of research: ______________________________

2. Is your present work primarily
   applied research aiming at solutions to practical problems _____ OR *basic* research aimed at creating new knowledge _____

3. How well do you feel you are able to keep up with advances in your field
   Very well _____
   Satisfactorily _____
   Not well at all _____

4. Estimate the relative importance to you of the following sources of information:
   a. professional literature (including duplicated material, etc.) _________%
   b. conversations and correspondence with colleagues, meetings, courses, and study tours _________%

5. Estimate the time you spend per week on searching and reading literature _____ hours/week.

6. How many scientific and technical journals and series do you personally
   a. subscribe to? _____ journals
   b. obtain through memberships? _____ journals
   c. follow regularly in addition to the above mentioned? _____ journals

7. Which of the following types of publications do you use?
   Mark those *frequently* used with 2
   occasionally used with 1
   scarcely ever used with 0
   _____ Journals and scientific series
   _____ Abstracts and indexes
   _____ Reviews
   _____ Books (monographs, handbooks, etc.)
   _____ Patents and specifications
   _____ Research reports (as individual publications, duplicated, etc.)
   _____ Other publications, namely ______________________________

8. List the foreign languages that you can read: ______________________________
   ____

9. Do you keep a personal index? yes _____ no _____

10. Do you have an opportunity to use the reference and information services of a library? yes _____ no _____

11. Which of the services listed below have you made use of?
    Frequently = 2, *sometimes* = 1, *never* = 0
    a. _____ Loans and photocopies supplied to you

b. ____ Quick reference service (e.g., numerical data, trade name given over the phone)
c. ____ Brief literature search on request (e.g., a selection of references of literature on method X)
d. ____ Comprehensive literature search on request (e.g., patent search or a bibliography of all the literature on method X published during a period of 2 years)
e. ____ Critical survey of the literature searched on your request in which attention is paid to the value of the reported results
f. ____ Continuous scanning of the literature on a subject requested by you and the "feeding" of abstracts or references or the material as it appears
g. ____ Translations of foreign language material into your mother tongue
h. ____ Abstracting articles specified by you
i. ____ Translation of your manuscripts into a world language and checking language before publication
j. ____ Editorial assistance, proof reading, diagrams, etc.
k. ____ Guidance by the library staff when you are searching the literature.

12. If you have not had an opportunity to avail yourself of some service listed under 11 which you feel a great need for, mark it with an asterisk. (Note: only *one* item.)

13. How many papers, research reports and books have you published during the past three years? ____ papers
In which journals have your papers appeared: ____________________

____

14. What are some of the difficulties you have had with your own publications? (such as lack of suitable journals, slowness in publication, lack of financial support for printing and translating): ____________________

____

15. What are some of the difficulties you have in obtaining required information and keeping up with advances in your field (such as time limitations, lack of access to published material, slowness of publications, and lack of suitable abstract journals)? Please specify: ____________________

____

____

16. Can you describe specific instances of undesirable duplication in your work caused by the lack of information on research carried out elsewhere? ____________

____

____

17. Can you recommend improvements which might make the literature and reference services meet your needs more adequately? ____________________

____

____

____

18. Have you after graduation felt the need to improve your skill in the use of literature? yes ____ no ____

19. Remarks:

The Transmission of Scientific Information: A User's Analysis

J.D.BERNAL

As a contribution to the present Conference this paper may seem out of place. It does not set out the results of any enquiry on the uses of scientific information nor does it confine its interest to the specific theme of the Conference on the *Storage and Retrieval of Scientific Information.* I understand very well why the Conference had to be so limited, for the subject of scientific information as a whole, which was dealt with at the Royal Society Conference ten years ago, has now grown so large that it can be attacked only piecemeal. Nevertheless, I hope that my contribution may still be acceptable because what I have tried to do is to present, from the point of view of a user of scientific information, the particular aspects of storage and retrieval (or what might be called the *memory* function of scientific information) on the background of the whole problem of communication between scientists, technologists, and the interested public.

It is perfectly admissible, indeed often necessary, to concentrate on the problems of storage and retrieval as forming part of a closed field. We may take the pieces of scientific information to be put into and taken out of store as simply *given* by the working of the rest of the scientific machine. The problem then is simply how to handle them most quickly, accurately, and economically by human or mechanical methods. This, no doubt, will be done in the fourth, fifth, and sixth sections of the Conference. There is, however, a danger, and necessarily a growing one, that the service of scientific information will develop as an activity entirely in its own right, ever more and more competent to take in, store, and hand out information regardless as to whether this information is superfluous, inaccurate, or unwanted. Fortunately the organisers of the Conference did stipulate that this activity had as its ultimate aim the service of the user and, therefore, it is legitimate to ask not only *how* the information is to be dealt with but also *what* that information is, *who* is it intended for, and *to what degree* does the process of transmission of information help in the advancement and use of science.

J.D.BERNAL Professor of Physics, Birbeck College, University of London, London.

Now I must admit at the outset that my views on this matter may be biassed because although I have been definitely a user of scientific information for the better part of forty years, I am not a representative user—though there may be no such creature—and not even a model user of scientific information. For it is evident that those engaged largely in fundamental research represent a small minority, possibly as small as five per cent, of the *users,* though, owing to the fact that they are obliged to read more individually, they may represent as much as twenty per cent of the *actual uses* of scientific information. Later I will propose statistical enquiries on the composition of the population of *users* and *uses* of scientific information. (See also Appendix.)

Even if my contribution is taken as representative of fundamental research users—and strictly it can be so only for crystallographers and some allied branches of chemistry, physics, and biology—it needs must express personal opinions and not objective statistical facts about such users. Here at the outset I would defend this position because I consider for various reasons, some of which I will touch on later, that there are fields of enquiry of such complexity, variability, and novelty that verbal and qualitative analysis should precede numerical analysis whether objective or by questionnaire.

I am here proposing only to indicate the nature of the enquiry, not to answer questions that might be raised in it. My concrete contribution will be to propose a series of enquiries all tending to elicit the user aspects of the problems of storage and retrieval of scientific information. (These will be described as they arise in the text and listed together in the Appendix.)

My main reason for presenting this paper at such a Conference is that I believe that the whole subject of transmission of scientific information needs an analysis of a descriptive or *natural historical* kind before we can hope to find the right figures to look for or the right questions to ask. This is not only to ensure that the answers we get are significant, statistically or otherwise, but also to determine whether the answers that prove to be significant and true are really relevant to the total situation that we hope to understand and control, namely an improved flow of scientific information.

The main reason behind this implied criticism is that if the matter be treated as one of operational research, it follows that all enquiries as to present uses of scientific information services, though a necessary background, can by themselves tell us little of use for improving the service. They tell us what people do with an admittedly very imperfect service, not what they would do with a better one (which would naturally include proper training for its users). A certain amount could be learned by a comparison between different systems in use, and some lessons from this quarter may emerge from our Conference, but we cannot hope to learn much until it becomes possible to carry out trials involving considerable variations under strictly comparable conditions.

The essential difficulty is that, though the user may well know what he *wants* from an information service, he is in no position to know what he *needs* from it, namely what variation in the system would help most to further his work. Consequently, any action based on analysis of present user habits is unlikely to produce impressive results.

What I will try to do in this paper is first to give my own natural historical account of the whole process of the transmission of scientific information, indicating in the course of this what further enquiries or actions, such as analysis, would seem to be called for. I do not attach as much importance to this latter aspect as I do to the former as I am more concerned with stimulating discussion than with the carrying out of actions which will in any case require much deeper consideration than I can give them before they are likely to be useful.

It is difficult, especially in these days, to visualise or describe the process of transmission of scientific information as a whole. I tend to see it as a complicated irrigation system which is continually fed from many sources and in which the individual plants (the users) depend on what reaches them at any given time. Ideally, each should receive just the right amount of water at the right time but in practice, owing to the sluggishness and irregularity of the system, some never reaches the plants in time, and much of it evaporates or runs into the ground on the way. At other times the flow of water is so abundant that the plants are waterlogged and cannot absorb what they need. The simile is too crude, for it misses two essential features of scientific communication: first, that the receivers are in their turn sources of information, and secondly, that it is not generalised but highly specialised information that is wanted. However, it may serve to bring out two related defects of the present communication system, its viscosity and its wastage. In general the path from the provider of a piece of information to any one of his recipients is so long that the latter gets it too late to obtain the full value from it, often too late to be of any value at all, for scientific information is a particularly perishable commodity; further, the longer the time the greater the chance that he does not get it at all. This is the factor of relative wastage of information which prevents *everybody* capable of profiting from a piece of information from getting it. More serious, though also still unmeasured, is absolute wastage in which *nobody* capable of profiting from a piece of information gets it.

If this were all, then science as a whole would gradually settle down to a slow rate of progress determined by the amount of information that managed to get through. Actually science is advancing very rapidly because there are other sources of information which can be found all along the line from examining nature itself by experiment. A working scientist or technologist needs information. He has the choice of getting that information through information services, or of finding it by experiment, or of working it out for himself. In that way the same fact, or method, more or less, may be rediscovered many times. The detailed history of science is so inadequate that we have no measure of how often this occurs, but that it does occur I am certain, for I have myself both discovered experimentally several things already known and had my own published work rediscovered experimentally by others (Appendix, Item 10).

The reason I stress this waste of effort and knowledge now is because in default of any reform of the system of communication, it is bound to grow with the growth of science. The sluggishness of the system is, however, in my opinion, more serious. Information, even vitally needed information, takes months or years to reach those most in need of it. We should reflect on our common experience when visiting our colleagues' laboratories, even in the same country. It is rarely that we do not learn something of importance to us that we did not know and as often impart a useful piece of information to our hosts; and this is usually already published information. In principle, therefore, neither of us has an excuse; if we had read everything, we would have known the facts already. But now nobody does read everything, and indeed nobody could, even if he did nothing else. The basic fact remains, *the amount to be read increases exponentially, and the time anyone has for reading it remains the same; therefore a smaller proportion of what is written is read by any one person.*

One escape from this is through increasing specialisation. Reading surveys show (1) that one scientist can keep up with the work of some two hundred others in active production, but a field which contains only two hundred workers is necessarily a very narrow one. Now this may not matter if the field in question is a newly developed break-through in the front of ignorance such as “subgenetic analysis of bacterial viruses,” but these can represent only a small fraction of specialist fields which are far more often pedestrian collections of knowledge or skills. Here what science loses by such enforced specialisation is the cross fertilisation of ideas from different fields such as lead to all great discoveries.

Another way round the difficulty is to limit reading not so much by specialised fields as arbitrarily by reading only a small selection of journals. Dr. Urquhart (2) has shown that out of 9100 periodicals taken by the Science Library in London, 4300 were not consulted at all in a given year. Now it is difficult to believe that nothing of interest to the 87,000 readers at the Science Library was to be found in these 4300 periodicals. If so, the sooner they cease publication the better. Rather it would seem that these journals were simply unfamiliar to the readers and inadequately abstracted.

Still another way is to recognise frankly that not everything relevant can be read and to dip into the literature almost at random, qualified by hints given by colleagues or information officers. Thus at least the reader can hope to get a statistical view of the current state of knowledge.

Finally there is the wise and instructed reader (I wonder what proportion of scientists fall into this category) who reads a half-dozen journals regularly, another half-dozen annual or quarterly reports in different fields, and keeps up with the rest of scientific information through the systematic reading and indexing of abstracts (Appendix, Item 6). It is just now that this last task is itself becoming too time-consuming and forcing even the conscientious reader towards specialist or eclectic courses.

Now to me it seems evident that it is here that the information services should step in and so process the raw information that each user will receive just so much and no more than he can cope with. It should at the same time give him the necessary breadth of information and answer the specific questions that he has been able to formulate, even if only vaguely.

Before we can see how this might best be done, I feel it would be worth while looking more closely both into the *users* and the *uses* of scientific and technical information. I must include technical information here because nowadays no scientist can do his work well without technical information and vice versa. Now from the point of view of the kind of information services required, users can be divided into the following categories which are strictly functional, for the same man can at different times belong to any of these categories:

(a) Workers in fundamental research.
(b) Workers in applied research or development, including medical and agricultural research.
(c) Technologists, including engineers, architects, medical practitioners, and agriculturalists.
(d) Writers of reports, textbooks, teachers, students, etc.
(e) Scientific and technical journalists.
(f) The interested public.
(g) Historians of science.

These categories differ very much in their relative numbers and in the degree to which they make use of information services. It should be one of the first tasks of a systematic study of such services to get some estimates of these quantities as a whole, broken down into subject fields (Appendix, Items 1 and 2). I should be prepared to guess that categories *(b)* and *(c)* are predominant, and of these *(b)* will be the most important for, though there are far more technologists, in the sense given above, than workers in applied research, the latter need to use far more information. I should suspect, however, that though the information wanted by technologists is less in total quantity, it is of a kind more difficult to find—where it exists at all—and gives a disproportionate amount of work to the retrieval service.

Category *(a),* the fundamental research workers, probably make the most use of information, but they are relatively few in number and, as I have already estimated, probably account for less than twenty per cent of the load on the service. Category *(d),* which includes most academic scientists at some stages of their careers, is not large numerically, but their special requirement for thorough coverage of a particular field subjects information services to a serious and probably salutary strain. However, all such workers are not merely users of the information services but also contributors to it—they should effectively give back to it as much, or more, than they take out. Categories *(e)* and *(f)* require more personal attention than mechanical retrieving services. As to *(g),* historians of science, retrieval of information is part of their professional skill. I believe, moreover, that our information services and particularly our libraries effectively provide them with far more material than they need or indeed want. For, as the Urquhart (2) study shows, the demand for periodicals more than ten years old is small, and that for more than thirty years old, negligible. Consequently, much would be gained by clearing shelves of these and leaving them to central libraries or archives.

The question of the *users* of scientific information must be distinguished from that of the *uses* which are made or hoped to be made of this information when obtained. This question is an extremely complex one and can be approached in a number of different ways, each of which has some bearing on the performance of information services. We may distinguish between an analysis based on the nature of the material used and one based on the activity of the user.

In the first mode we can roughly divide the information sought into *(a)* data to be used in practice or incorporated in research; *(b)* procedures, techniques and methods, including descriptions of apparatus; *(c)* conceptual frameworks, theories or ideas—these are not only presented to be used but also to provide inspiration: positively by extending them or negatively by criticising them. To this may be added the most elusive category of transmitted information, namely absence of information, the suggestion which does not arise of itself, of gaps in knowledge, often the most fruitful impetus to new work.

Now it is evident that these categories offer very different problems to the information transmission mechanism. The data are far the simplest, and it is natural to think of them as the only material to be handled by information services. Here all that seems to be required is that they should be as *accurate,* as *fresh,* and as *complete* as possible. Data, indeed, are the only types of information which can be safely detached from their original sources. The transmission of techniques is much more difficult and deserves very special consideration. It is a field in which abstracts and reports are almost necessarily inadequate. At the moment the difficulty and slowness in the transmission of techniques furnishes in itself adequate justification for visits and exchanges of scientists between laboratories.

It is the last category, that of theories and ideas, which is the most difficult of all to deal with by any means other than through original papers or personal contact. Few people are capable of understanding a genuinely new idea or theory. Very few, possibly only one or two, are capable of profiting by it. Here the essential function of the information service is to see that these people hear about it. This is where reports are more valuable than abstracts. Naturally, this difficulty does not occur at the growing points of science where there is the greatest proliferation of new ideas and all the workers are alive to their possibilities and anxious to catch at them. It is rather at the boundaries and dead ends of fields of research that ideas are apt to be lost, especially as the first germ of a new theory is bound to be, or at least seems to be, of a crazy character.

If we turn now to the second mode, the activities of the user in his capacity of wanting information, we may break this down first into *general* and *special* uses. The first is the task of keeping up with knowledge in general over the whole field of science, with a greater and greater concentration on the particular field of interest. This need is met for the outer field by the service of books and science magazines, less or more specialised, and in the central field, by half a dozen or less journals. The service is by no means perfect, but short of reorganising the whole of scientific publications, which will have to be done some time, there is little that can be done about it now by the information services as such. Indeed, the very chaos of present-day publications and the unpredictability of their contents automatically ensures a more or less random sampling of the field of science by the average scientific user. This might be systematised by a reader picking and reading through one paper in ten in his own field, one in a hundred in neighbouring fields, and one in ten thousand in the rest of science, at least of those parts whose language he could read. I doubt whether for most people this would be any better than the casual picking up of papers that goes on now. Far more important would be an improved, graded, up-to-date and not too abundant service of reports of progress in the various branches of science.

The scientific and technical user, besides these general or background requirements, has need of specific pieces of information. This need has been the raison d'être of special information services beyond the scope of the distribution of periodicals and the functions of libraries. This special information the user hopes to incorporate in his own work, or at least to have the chance of so incorporating it. For the same reasons he wants it quickly, and he wants to be able to rely on it or at least to know the evidence by which he is expected to rely on it.

It is at this point that the needs of different kinds of users diverge, though there is still much overlapping. The fundamental research worker and, to a lesser degree, the applied research worker, require scientific information as a *central* and *essential* part of their work. Their need for it, indeed, is greater than for the new results which they are themselves getting out. For inevitably, except in completely new fields, all they can hope to add is a small part to an already considerable edifice.

However, this problem of acquiring what might be called *central* information is not one which calls much on information services except indirectly. For the research worker would not, in general, be doing the work he does unless he already had, by his own efforts or through those working with him in the same project, mastered existing sources of information in his field and knew where to look for new results as they come out.

The exception is when he is entering the field for the first time, not so much for the student in his first research, where he can usually count on guidance, but for the mature worker who wants to branch out into a new line. This is fairly infrequent in fundamental research but extremely common in applied research where projects are often embarked on in fields quite unfamiliar to the research teams. The lack of provision for what might be called *pilot* information is one which I have often felt myself. I am unfamiliar with any steps that are taken to meet it, but I imagine that something of the kind is a regular practice in the larger research and development departments of industry and government.

It is probable that the main call on information services—though this point deserves to be tested—is not in respect of pieces of information that are central to the work of the enquirer but rather *ancillary* to it. The working technologist is generally occupied in work for which he already has the essential knowledge, whereas the research worker, as already indicated, has already his own means of finding the information in his special field of study.

Where technicians and scientists alike need information is in aspects of their work in fields with which they are not necessarily familiar. This is particularly true of apparatus, methods, and the properties of substances, where advances in one field could be utilised by workers in others as soon as they can find out about them. Here the initial difficulty is that the enquirer is likely to know only vaguely what he wants, and most of those who need the information do not even know it exists and consequently never get as far as enquiring. The limited success of Dr. Urquhart's “Unanswered Questions” scheme might make it appear that the need here is not great. However, I am convinced, particularly through my visits to laboratories, that here is a case where real *needs* greatly exceed felt *wants*.

Some years ago the British National Physics Laboratory sent round some of its expert staff to laboratories of research associations covering different aspects of applied science, such as rubber, flour milling, pottery, and leather. They found that, while in their own fields the knowledge of equipment in these establishments was well ahead, this was not the case where it lay outside their special field of competence. Here, equipment was often devised with great ingenuity to answer questions as they arose without the knowledge that similar equipment had already been developed in some quite different field of science. Given adequate information, it could have been used with slight adaptations and saved much trouble and time.

Cases of this sort must be far more frequent than we know and emphasise the necessity, particularly in applied science, for a *positive* information service that provides needed information unasked. I know that such services are provided in some of the larger research laboratories, but what we should seek is to spread them over the whole field, particularly to the smaller, two- or three-man laboratories so common in industry and in the fields of agriculture and public health.

So far in discussing users and uses of scientific and technical information I have treated all kinds of scientific information indifferently, without reference to subject divisions. This was necessary in order to compress this study within manageable bounds. However, I am fully aware that in any practical scheme for improving the flow of scientific information, the adaptation of the method to the field is perhaps the first requirement. Indeed, as information services have grown up, over the years, largely independently in different fields, they have developed different and even divergent methods of dealing with their problems.

I realise, for instance, that my own experiences, lying in the field of the mathematical-mechanical-physical sciences where much depends on the discovery and application of relatively few *principles,* are biased and do not adequately take account of the needs of the biological-geological descriptive sciences where the problem is to find the relations of vast numbers of originally unconnected *facts*. The problems of storage and retrieval are probably much greater in these fields. This is not only on account of their extensiveness in material but also on account of their much wider time range.

Many studies (3), of which Dr. Urquhart's (2) is probably the largest, show that the effective life of a piece of scientific information in the different fields of science is vastly different. The *true half-life* of a particular piece of information can be defined as the time after publication up to which half the uses (references) or enquiries about the piece of information are made. This is naturally extremely difficult to evaluate, though it would be well worth doing. Instead we are obliged to use what might be called the *back half-life* of a group of similar pieces of information—papers in a given journal for instance. This can be defined as the time counted back from a given date within which half the requests for, or references to, information have occurred. This period is about two years for physics and fifteen for biology.

This in itself indicates that the storage and retrieval apparatus in biology will always have to be much greater than in physics. It also indicates that it is quite possible in view of the ephemeral nature of information in physics that there would be some room here for economy in storage, and this might balance the need for greater speed in retrieval. In general my plea here is that in remodelling storage and retrieval, no effort should be made to achieve uniformity, but rather that a set of interlocked systems should be perfected by making full use of experience in each field.

Before discussing the second of my general questions as to the nature of the scientific information to be stored and retrieved, I must say something more about the scientific *user* in his other capacity as a *producer* of information. Now this aspect, as such, lies outside the scope of the conference, but it should not, in my opinion, be altogether excluded. We may be prepared to accept—but only for the purpose of these discussions—that there is just nothing to be done about the growing chaos of scientific publication. Our business here is to take the results as we find them, if we can find them, put them in the best order for storage, and hand them out where they are wanted.

However, I think that, even for this limited purpose, it is necessary to look a little closer into the production and fate of the material that is being so handled. The writer of a scientific or technical paper is trying to fulfil a duty to science and, incidentally, to establish his own reputation, by making his results known to the scientific world. This is at least his hope and belief, but it is one which it is increasingly difficult to realise, basically because there are so many thousands of other scientists doing the same thing. I pass over here the abuses of unnecessary, inflated, and multiple publication indulged in for prestige or simply to secure jobs. Certainly everyone, even former offenders, would gain if all scientific publications were genuinely new, concise, and accurate accounts of work done. It would also be possible for the reader to cover more ground if the length of normal scientific papers was shortened in something of the way practiced for centuries in the *Comptes rendus* of the French Academy or along the lines of the proposals of the Royal Society's Information Conference (3). It might be possible by such reforms to prune the material to be read by as much as half or three-quarters, but as it normally doubles every ten years or less, these steps, however welcome, would be temporary palliatives. The central difficulty, the growing *plethora* of scientific publication, would remain. It has already gone a long way to bury new work under the mass of print which there is no time to read.

The average scientific author of today—I am not talking of men of established reputations or easily recognised young geniuses—has less chance of having his work understood and made use of than at any previous period in the history of science. There is just not time to read all the papers, and the chance that any particular one will find the reader or readers who will make use of it within the very short effective life of scientific publications is small; exactly how small, it would be very interesting to know (Appendix, Item 9). Many knowledgeable people would put that chance very low; indeed few would suppose it to be a certainty. I have myself suffered more than once for publishing papers twenty years before their time and having them justly ignored or misunderstood.

Now this brings us to an illuminating remark of the organisers of this conference which might mark a great departure in the handling of scientific information. In the paper on Conference plans (Area 2) we find *"The primary reason for publishing (original scientific work) is to disseminate scientific information, not to store it."* I was very struck with this statement, and for all that it may appear to some as a truism, I feel that it was highly significant. However, the conclusion I draw from it, which does not seem to be quite that intended, is that what is revealed here is a contradiction in terms. *It is in present conditions growing more and more difficult, and may soon be impossible, to disseminate scientific information unless this is done in a fashion that permits its easy storage—or at least its easy processing through a storage and retrieval mechanism.*

There was a time when scientific information was not *disseminated*—it was *communicated* by word of mouth or letter to those deemed by the original writer likely to be interested in it. Dissemination through scientific journals was an indication of a scientific world which had extended beyond the sphere of personal acquaintance. Now in its turn dissemination, or in English *broadcast scattering,* is failing on account of a still vaster growth of the scientific public. To use a biological simile, the method of transmission of scientific information is on the most primitive level of wind-blown pollen. The more pollen —above a certain level—the more miss the stigmata waiting to catch them. We ought to advance at least to the more selective stage of insect-borne pollination where with far less pollen more gets to the right flowers. What we need now is to return to scientific *communication* on a vastly greater numerical scale and to make use of complex organisations and mechanisms in place of inefficient *dissemination*.

The contradiction I cited above is only an apparent one; its resolution can be found by treating original scientific publications no longer as the main exchangeable currency of scientific information but as the raw material for processing storable and retrievable information for the general reader. This implies in no way a restriction of the amount of direct contact between the original worker and his reader. Papers would be written as before and would be read in their original form by as many, if not more, readers than at present, that is, if the secondary publications—abstracts and reports—can be improved in coverage and speed. The only difference would be that the people benefiting by this information would no longer be limited to this circle. By passing all original material through the appropriate processes of classification and condensation, it would be possible to make at least their *factual content* also available to a much larger set of users. Further, by the use of positive information services it should be possible to ensure that at least a perceptible fraction of those that should benefit from the new information should hear of it.

There is an additional reason, quite apart from the compelling one of crowding of papers, for extending the range of the original publication. It seems to be a fact that the practical technological reader finds original scientific papers and even original technological papers very difficult to read and does not read them much. I can justify this statement only by hearsay, and a special enquiry ought to be useful here (Appendix, Item 5). It might also find, if the statement is true, why he does not read them or thinks he cannot understand them. It would seem, therefore, that *those who have the greatest need to follow the advances of science have the least opportunity to do so*. This situation could be remedied by adapting the results of original investigations to the interests of the practical users. This is a task undertaken today by the technical press and possibly it is as well done in detail as it can be. However, the problem of plethora is now beginning to be felt not only with original papers but also with specially written articles. Indeed, uncontrolled multiplication of scientific information, when the simple information itself is too much to get round, is, however well-intended, self-defeating.

So far I have indicated only the need for the processing of scientific information. Now I should like to put forward some tentative ideas as to how it might be done, or at least set out some conditions that any such schemes ought in my opinion conform to, if they are to be of use and not merely one further addition to the mass of secondary scientific literature. What I have in mind goes somewhat beyond what has hitherto been considered the function of information services, although for some thirty years it has been done in the field I know best, that of crystallography. In principle its purpose is to supplement the present mechanisms of storage and retrieval through abstracts and indices which treat the object to be retrieved as the *original paper,* by another system of reports which takes this object to be the *facts* contained in a number of papers.

Casting back to earlier remarks on the various uses made of scientific information, I distinguished three main categories: *(a)* data, *(b)* procedures and methods, *(c)* conceptual framework, theories, and ideas. Only the first of these, *data,* is at present suited to simple storage and retrieval techniques, and over many fields these are being so stored, while better and better methods are being designed for their complete and rapid retrieval.

The problem of storing *procedures* and *methods* including descriptions of equipment is intrinsically much more difficult, but it is one which would be of enormous value to solve, for I am convinced that the present slow rate of transmission of techniques, particularly instrumental techniques, forms at the moment the most restrictive single factor to scientific advance.

The third problem, that of dealing with *ideas,* is the most difficult of all, indeed it might be argued that it is intrinsically insoluble. The reasoning by which such a conclusion can be reached was given by a distinguished mathematician who, at one of the Royal Society's Conference discussions, objected to any form of abstracting. His contention was “No one can abstract my papers; all they could show by trying to is that they did not understand them.” “Then why not abstract them yourself?” “Impossible; if they could have been written in a shorter space I would have done so.” “Well then, if they are not abstracted they will be read by no one but your friends.” “These are the only people I write them for.”

Without going so far as this, I would consider that we are still so far from being able to abstract and reproduce in classifiable and retrievable form this aspect of scientific achievement, that at this stage it is not worth trying to do so. Further, as nearly every scientific paper contains some ideas or theories new, at least to the author, this in itself furnishes a justification for preserving the access to all original papers through a complete abstracting and indexing system. In other words, this *direct channel of transmission must be kept open,* for even if only a small proportion of the total of scientific information flows through it, this amount is essential to the maintenance of the growth of science.

To admit that, whatever changes are made, the established means of access to original work is preserved does not mean that I would grant that the present methods of publishing and disseminating information in that form are ideal. I still feel, for reasons already given, that instead of the present intermediate length paper often to twenty pages, it would be better to have a short, pointed paper of some two pages in the form of what has been called an informative abstract. This would be supplemented by a longer, more detailed paper, not printed and published, but available in duplicated, microfilm, or other modern method of reproduction, to all those thought to be interested in it or who requested it.

Such proposals, however, are outside the scope of the present Conference; here I want only to discuss the subject of secondary publication covered in Area 2. It seems to me that for this purpose we can divide the field into two aspects of the recording of science, what might be called the *differential* or *current,* and the *integral* or *cumulative,* both of which need to be served by different forms of secondary publication.

The first of these aims at bringing to the user an ordered picture of the present activities of science. This would be less immediate than the existing type of science magazines like *Nature,* but more so than the normal reports of *Progress in* ______ which are proliferating today, though like them it should carry enough indication to find the original literature. A time delay of six months from the latest entry, to eighteen for the earliest, is, I think, technically achievable. It should be recognised, more than it has been, that these kinds of reports, like the original papers on which they are based, are *ephemeral publications*. They would be aimed, on one hand, at giving notice of the working hypotheses and new discoveries of the day, but in no sense presenting a record either for reference beyond five years or so, or for history which requires quite a different technique. They must also be written by active and interested workers in the field, taking precautions to avoid or neutralise personal bias.

The *cumulative* secondary publications by contrast will not be intended for reading, but for reference. Already, over all the easily reduced parts of science they exist as Mathematical Tables, Data Tables, Floras, and Handbucher, etc. There is little that need be said about these except that they need more money and help to keep up to date. Further, as no doubt will be fully discussed at this Conference, they need to be adapted to modern methods of search and retrieval, even to the extent of superseding the old printed volumes.

There remains, however, the much more serious problem of extending the ideas of cumulation of *data* or *facts* in fields where hitherto there has only been accumulation of *publications*. We have to find out whether this is mainly because there has been no demand for such data or merely because the problem of collection of such information and its reduction to standard forms suitable for data compilation has proved too difficult. These two aspects are related. The real need for data compilation cannot be fully felt as a want until there is some possibility of meeting it.

The two great difficulties that have held up the extension of data collections to other fields of science have been those of c*lassification,* which grows with the complexity of the particular system of knowledge, and of *reliability* of the data themselves, which varies both with the difficulty of the subject and the rapidity with which it is advancing. I would think for instance, though I have no competence in this field, that in many branches of medicine standardised and statistical information would be of considerable value, though I can see how difficult it would be to prepare them and keep them up to date.

In another field where I do have some experience, that of scientific instruments, I believe that collections of data would be of value. I ground this belief on my own experience in visiting laboratories and on that of the National Physical Laboratory, already referred to. Here the difficulty is not that of complexity nor of classification, but of speed of advance and consequent obsolescence. The old method of printing data tables or dictionaries would here be quite useless, except for the history of science, because the information is wanted within a period reckoned in months rather than decades. However, I believe it would be an admirable field to try out rapid electronic sorting and retrieval devices once a sound and flexible method of classification had been worked out. Data on scientific instruments have one advantage not shared by many other branches of science. Because an instrument has to work and have a measurable performance, the original data are largely self-checking. In other words it would here be relatively easy to set up a semi-automatic filter against repetitive, irrelevant, and incorrect information.

The inability to do this is the great curse of all data compilation, for experience shows that the inclusion of even a small proportion of bad information in any set of tables results in a feeling of untrustworthiness that spreads over the whole collection.

Data collections are too often considered only in their passive aspect but, except as the subject of historical studies, they need continual working over, of which they never get enough, as the bodies responsible for them are generally so short of means that they have all they can do to get the new data. Good data tables need to be reclassified and rearranged at decennial intervals, or more frequently in rapidly moving fields. The switch over from printing or card methods to newer techniques should make such checkings and rearrangements not the annoying interruptions of routine they are now, but a normal part of the process of information storage, in ways of which we may hope to hear much during this Conference.

Such tables would have a value beyond their use for reference. They should suggest by their mere systematic arrangement, by the use of statistical or other methods, new generalisations and laws in the fields they cover.

As time goes on, we might expect a change in balance between the uses of the two types of secondary publications already discussed, that is, between that based on the storage and retrieval of *primary publications* by abstract and index systems and that based on the current and cumulative handling of *facts* given by reports and data compilations. As the former become more and more unmanageable by mere bulk, the latter will, in my opinion, gain in importance. It would be interesting to enquire to what degree this is already occurring (Appendix, Item 7).

I have now covered the main objectives of this paper in presenting an analysis of scientific information services from the point of view of the user. I must apologise for its qualitative and personal character, but it may for that very reason form a useful corrective to the predominantly quantitative and mechanical character of the Conference. The object we all have in mind is not that of processing a material or even a set of figures, but rather of effecting the largest measure of communication between human minds.

Before I conclude, however, I would like to add certain remarks and suggestions referring to the areas of the Conference concerned with enquiries on the use of information and on methods of search and retrieval.

I am a strong believer in the value of operational analysis in the field of information transfer. Indeed, I had a striking example of this in the conclusions from the rapid pilot survey of reading habits carried out for the Royal Society Information Conference (3). Before that, I had been so much impressed, through the experience of my own work, with the importance of reprints that I had proposed a scheme for substituting a rational distribution of these for the traditional scientific periodicals. This scheme roused much feeling and was even castigated in a *Times* leader as “Professor Bernal's insidious and cavalier proposals.” However, the result of the pilot survey showed me that scientists as a whole did not work the way I did, but rather made use primarily of libraries where the disadvantages of the bound periodicals largely disappeared. Consequently, I immediately abandoned my original proposals and publicly withdrew them at the Conference.

Here I might have been precipitate and wrong, for the fact that working scientists used libraries only proved that with the existing system of distribution this was the best they could do and that they might well change their habits to advantage in the new conditions. However, operational proof that this was not the case was furnished by the results of the use of a scheme, similar to mine, but more limited, by the Physical and Chemical Societies in Britain. Both schemes failed and for a reason I had not anticipated, namely that working scientists cannot be bothered even to make a mark on a form to get the paper they are interested in. Whether it would have worked in the way I originally suggested by which they would have received the papers in specified fields automatically remains to be seen, but I am inclined to doubt it.

Despite the proved value of such enquiries, backed by operational variations, I would like here to put in a very general plea against treating these enquiries on the level of ordinary scientific investigations and accepting the results at their face value. In all matters involving human behaviour, even that of professionally rational scientists, it is very difficult to infer from measured behaviour in one set of circumstances, even subject to some variation, what they would be under very different circumstances. In mathematical terms one method of measurement with slight variations may help to determine the local maximum of performance, but there may be other and much higher maxima which can only be reached by a considerable jump.

Applied to the problem of scientific communication, it means that we ought not to spend too much effort on routine studies of user behaviour except for pilot surveys to bring out striking but unrecognised features, such as that of Dr. Urquhart's library request analysis. We should rather try to find or create extreme conditions which should test our conceptions of the whole function of information. For instance, we might see how scientific research, undertaken without any information at all, would go in comparison with that using the whole battery of modern information services (Appendix).

My last suggestion is one which might be considered to run counter to the whole trend of interest of the Conference in favour of the mechanisation of information services. I do not think, however, that, in proposing a greater use of the experience and knowledge of individual, active scientists in the whole procedure of information services, I am in fact taking an opposition view, but rather proposing a very necessary complement to the increased use of mechanical methods. What I feel we need to do is to find the areas most suited to the capacities of the human mind and those most suited to mechanical processes.

If I may anticipate, for argument's sake, the results of such enquiry, it seems to me that the human mind excels in judgment and in search. As to the former I would use people, preferably several of them to avoid bias, in the first or filtering stage of processing of information in order to remove at the start the repetition and nonsense that is in danger of choking scientific information to death. This I would suggest would be best done by associating groups of competent and, for the most part, young scientists with the first stage of examination of the literature to extract the facts or alleged facts from it. This is indeed already an old practice and is used on a large scale in the preparation of the abstracts for abstract publication of the Soviet Academies (1). Every paper and its abstracts are read by three such auxiliary abstractors.

If the service were properly organised and sufficiently widespread, the load on individual scientific workers would be small. Where it might bear most heavily, in the preparation of reports, it would be of value to the generally younger scientists on whom it would fall. There is no better way of learning a subject than in preparing reports on it, as I have experienced in the years (1) in which I helped to prepare the *Annual Reports* of the Chemical Society. There is also no better way of appreciating the shortcomings of the scientific literature and the working of scientific information services and acquiring a desire to see them reformed. The association of all junior scientists (not just a selected few) in documented work in this positive and useful way would, in my opinion, be a most valuable introduction to them of modern documentary science. Many of the enquiries reported on or proposed at this Conference would give very different results if the scientists whose behaviour is being studied had a thorough grounding in the use of libraries and information services.

It is, however, in the other capacity of search that the greatest potential and least used capacity of the human mind lie. Being evolved largely for the selection of significant material and with a memory adapted for the purpose, the human mind skips over the field of choice and arrives at the desired point —if it gets there at all—quicker than a machine of many times the intrinsic speed which must work exhaustively.

But this is only the least important of its services in the matter of search. A machine can in principle find everything that has been written about a specified subject. For lack of judgment it may find too much, and most of what it finds may be nonsense, but that could, if necessary, be sorted out by the recipient. What the machine cannot do, however, and the human mind can, is to answer questions that have not been asked. The first sensible answer to any enquiry, except to the few who really know precisely what they want, is to ask in turn “Why do you want to know that?” If the questioner persists and swallows the insult, it is usually possible to give him what he really needs to know, rather than what he thinks he wants to know, and this all the more because such enquiries are usually made in fields so unfamiliar to the questioner that he literally does not know what to ask for.

Now I know that much of this function of eliciting information needs is already performed by the officers of the larger information services today. My plea is for an extension of this to make use particularly of the experience of older scientists who do not perhaps always know the answers or even where the answers will be found, but who do know the man who will certainly put them on the track. I remember well during the war, when at a Conference to unify military information services, the Chairman, the Librarian of the Admiralty himself, showed us the futility of our efforts by remarking that in all his forty years of experience he had never found that anyone had failed to get the information he wanted, provided he knew the right person to ask! Without going quite as far as that I would like to add a human who's who selector to the battery of mechanical information searching devices.

And now in bringing this paper to an end, I shall recall for the benefit of those who have lost the thread in its discursiveness only the main point I wished to make, which is that a knowledge of the requirements of the different users of scientific information and the uses to which they wish to put the information they secure should be the ultimate determining factor in the designing of methods of storage and retrieval of scientific information.

REFERENCES

1. RALPH R.SHAW. *Pilot Study on the Use of Scientific Literature by Scientists,* Rutgers University, New Brunswick, New Jersey, 1956.
2. D.J.URQUHART, "Use of Scientific Periodicals," this conference.
3. Royal Society Scientific Information Conference, June-July, 1946. Published 1948.

APPENDIX LIST OF ENQUIRIES IN THE FIELD OF SCIENTIFIC INFORMATION PROPOSED IN THIS PAPER

1. Statistical estimates of the number of enquiries for scientific information in the categories of *(a)* fundamental research, *(b)* applied research and development, *(c)* technologists, *(d)* writers of reports, textbooks, teachers, students, etc., *(e)* scientific and technical journalists, *(f)* the interested public, *(g)* historians of science.
2. Statistical estimates of the number of enquiries emanating from the above categories of users. (It would be preferable to combine 1 and 2 and to add a further breakdown into subject fields.)
3. What proportion of information required is needed for purposes ancillary to the main work in which the user is engaged?
4. What proportion of information required here, not only in enquiries but in books and periodicals asked for, is for general and what for special purposes?
5. How far does the technological world read *(a)* original, *(b)* secondary articles?
6. What proportion of users of scientific information has any special knowledge of bibliographic techniques?
7. How has the reading of primary and secondary sources of information varied within recent years?
8. How do the true and back half-lives of original papers differ in the different subjects?
9. What are the chances of a scientific paper in different fields finding one reader who will make good use of it?
10. Can any measure be made of the proportion of facts in different sciences that are rediscovered?

In addition I have proposed a competition in scientific research in the same field of three teams *(a)* with best available information services, *(b)* with present average information services, *(c)* with no information services at all.

An Operations Research Study of the Dissemination of Scientific Information

MICHAEL H.HALBERT and RUSSELL L.ACKOFF

The research reported here is still in progress at Case Institute of Technology under the sponsorship of the Office of Scientific Information of the National Science Foundation. Between this writing and the International Conference on Scientific Information, work on this project will continue. Only preliminary results are available at this time. The additional results obtained in the next six months will be reported at the time of the International Conference.

GENERAL DESIGN OF THE RESEARCH

The question originally asked of the Operations Research Group at Case by the Office of Scientific Information was: What is the possibility of applying Operations Research to problems in the dissemination of recorded information? The research reported here is a partial answer to this question. To understand its development it is helpful to be aware of two essential characteristics of Operations Research. The first is that Operations Research is concerned with the application of scientific method to the study of *systems* of organized activity rather than to the components of such activity. Its orientation is "whole-istic." Secondly, Operations Research is *operationally* oriented. This means that its primary concern is with affecting the way systems operate and not merely in providing "interesting information." In brief, it seeks to provide a basis for effective action.

With this much in mind we can reconstruct the logical development of this project.

The system of scientific communication can be thought of as having three phases: (1) production, the formulation of a message which may take the form of an article, book, memorandum, speech, conversation, etc.; (2) distribution,

MICHAEL H.HALBERT and RUSSELL L.ACKOFF Operations Research Group, Case Institute of Technology, Cleveland, Ohio.

the dissemination of the message as by publication, presentation at a meeting, etc.; (3) consumption, listening to or reading the message.

Of these three types of activity the most organized is the phase of distribution since this involves the activities of scientific societies, research institutes, publishers, libraries, and related organizations. Production and consumption of information is primarily individualistic and independent activity which, because of its lack of organization, is the most difficult to manipulate or control in any way. Consequently, the attention of this research was concentrated on the distribution phase of the communication process, but this concentration involved a concern with the interaction of the other phases with distribution. That is, it was decided that improvements in scientific communication could best be obtained by manipulation of distribution, but this did not mean that production and consumption could be ignored. To the contrary, changes in distribution could only be evaluated in terms of their effect on production and consumption.

What aspects of distribution should be studied? Consideration of the activities of organizations engaged in the dissemination of scientific information led us to conclude that their primary problem involved the question of how their resources should be used. Their resources include not only money, but also men, material, and machines. That is, how should the resources of these organizations be allocated to the various ways of disseminating scientific information so as to maximize the effectiveness of the communication system?

To answer such a question a measure of effectiveness of the system is required. It seemed clear that the system is concerned with increasing *scientific productivity*. It seemed equally clear that an acceptable measure of scientific productivity was not likely to be obtained within the time available for this project. Nevertheless, it is of value from the viewpoint of research design to consider how an answer to the problem of resource allocation would be obtained if such a measure were available.

Ideally, we would first like to identify all the alternative ways in which the organizations considered could invest their resources. For example, they might increase their abstracting services; they might initiate digesting or reviewing services, or they might develop new cataloguing systems. These are but a few of the many alternative types of activities in which these organizations could engage. Ideally we would like to identify all of them.

Secondly, for each possible alternative we would like to determine how any specified investment could best be used. For example, if, say, $100,000 were to be made available for improving abstracts, how could this sum be used so as to maximize the resultant increase in scientific productivity. Put more generally, we would like to generate a “pay-off” function for investments of various amounts in each alternative activity, assuming these investments were used in the most effective way.

If the alternatives could be completely identified, and if the best use of an investment in each could be specified, and if a pay-off function could be derived for each alternative, then we could construct what is called an "allocation model" of the system. Methods are available, once such a model is constructed, for determining for any amount to be invested, how best to invest it; and, more generally, how much ought to be invested and how.

Such thinking, as we have indicated, is highly idealized, but it provides a standard against which a practical research design can be developed. We would like to come as close to the ideal as possible. The practical adjustments that are required arise primarily out of two difficulties. The first we have already mentioned: the difficulty of developing an acceptable measure of scientific productivity. The second difficulty arises out of the very large number of alternative courses of action which are available to disseminating organizations.

Since we could not expect to measure scientific productivity directly we sought an aspect of scientific activity which (1) could be measured objectively and (2) if increased, would also increase scientific productivity. The *time available for scientific research* is such an aspect of scientific activity.

An effort which attempts to maximize the amount of time available for scientific research has several advantages. First, it is widely recognized that we are suffering from a considerable shortage of scientific manpower. Consequently, any activity which promises to make more research time available with the same number of men may help to solve this shortage problem. Secondly, the amount of time available for scientific research can be measured objectively.

It should be noticed that, although there is an assumption implicit in the use of "time available for scientific research" as a measure of effectiveness to the effect that if this time is increased productivity will also increase, there is no assumption made about the nature of this relationship, that is, its mathematical structure. It does not follow from this assumption, for example, that doubling the time available for scientific research will double scientific productivity. All that is assumed is that there will be *some* increase in productivity.

Because of the large number of alternative courses of action that could be considered, it is necessary to have a preliminary screening which will reduce this number to a manageable magnitude. If we knew how scientists actually spend their time now, then we could concentrate on alternatives which are most likely to affect significant portions of that time. Consequently the practical research design was conceived of as addressing itself to three questions in sequence: (1) How do scientists actually spend their time? (2) In what types of scientific activity are there the greatest potentialities for reducing time expended without reducing scientific output? (3) How can these potential reductions be realized in the most effective way?

At the time of this writing (1) has been answered for a specific science and (2) is answered in part. The third question has not yet been considered. By the time of the conference it is hoped that (2) will have been completed and that significant progress in (3) will have been made.

HOW SCIENTISTS SPEND THEIR TIME

In order to study the way scientists spend their time it is necessary to have a classification of such expenditures. The classification is critical. For example, one might observe what portion of their time scientists spend wearing jackets, but this is not likely to indicate fruitful possibilities for saving time. We need a way of looking at time expenditures which suggest things we can do to affect these expenditures.

First, it was clear that we wanted to distinguish between time spent on scientific activity and time spent on other types of activity. It was also clear that we wanted to distinguish between time spent in communication and time spent in other activities. This yielded a basic fourfold classification. It is also necessary to know how time spent on non-communicative scientific activity is allocated. This type of activity seemed to be conveniently classifiable into: thinking or planning alone, setting up or maintaining equipment, using equipment to generate data, and treating data.

Nonscientific activity can be divided into business activity and that which is personal and social.

These considerations led to the following basic classification of activities of scientists:

Scientific communication
Non-scientific business communication
Thinking or planning alone
Equipment set-up and maintenance
Equipment use
Data treatment
Personal and social
None of these
Out of area

"None of these" is included to cover any miscellaneous activities that cannot otherwise be classified. Discussion of the observational plan in the sequel will show why "Out of area" was required. Detailed definitions of these categories will be found in Appendix 2.

Since our primary concern is with scientific communication, this category had to be further broken down. This breakdown is essentially concerned with *(a)* the phase of communication: sending, retransmitting, and receiving, *(b)* who is involved in the communication, and *(c)* the channel or medium of communication employed.

Space does not permit detailed justification for the classification of these three phases of communication, but, fortunately, the classification, we think, tends to justify itself. The phases of scientific communication were broken down as follows:

1. Hearing question
2. Reading question
3. Reading for use
4. Reading for general information
5. Hearing information
6. Working out material
7. Editing material received
8. Writing information
9. Telling information
10. Writing question
11. Asking question
12. General discussion
13. Discussion about a received communication
14. Reading for retransmittal
15. None of these

The number of the following types of persons involved in each communication was sought:

1. Mathematicians, statisticians
2. Physical scientists (and engineers other than chemical)
3. Chemists (and chemical engineers)
4. Biologists and medical men
5. Behavioral scientists
6. Secretaries, technicians, etc.
7. Other personnel in the same organization
8. None of these

The channels involved were classified as follows:

1. Oral
2. Unpublished written
3. Book
4. Article
5. Abstract or review
6. None of these

Explanations of each of these categories can also be found in Appendix 2. The observational forms themselves will be found in Appendix 1. But these will be better understood after a discussion of the remainder of the research design.

Several different methods of collecting information on how scientists spend their time were considered and tested in the field. An effort was made to develop a procedure which would yield reliable, accurate and objective data. These requirements dictated that the report made on the scientist should be prepared by someone other than the scientist himself. The classification of activities lends itself to use by an observer. Only a few categories required any consultation with the scientists themselves, and these involved only questions of fact, not a solicitation of opinions or attitudes. The pretests supported these contentions. In order to get a representative sample of time the "ratio delay" method was adapted to our purposes. This consists of observing the scientist at randomly selected moments of time. From observations so made estimates can be made of the over-all allocation of time and the accuracy of these estimates can also be determined.

After extensive pretesting and collection of information as to the time required in making observations and in setting up to make observations, and after estimating the number of observations required to obtain results of acceptable reliability, it became clear that the resources available for this study would permit analysis of only one scientific discipline. The field of chemistry was chosen primarily for two reasons:

1. Chemists constitute the largest scientific group, numbering approximately 80,000.
2. Among scientific groups the chemists have probably been among the most concerned with problems involving the dissemination of recorded scientific information.

It was considered essential to select the chemists to be observed in such a way that inferences could be drawn from the resultant data to as large a portion of the total population of chemists as possible. It was our feeling that in too many studies of scientific communication groups are selected for their convenience and consequently provide no scientific basis for generalization of results obtained. Consequently, with the cooperation of the American Chemical Society and the National Science Foundation a proportionate stratified systematic random sample was drawn from the population of chemists in the 150 metropolitan areas of the United States. Details of the sampling procedure are given in Appendix 3.

The sample yielded 50 observed groups of chemists, employed by 45 different industrial organizations and 5 groups of chemists in universities. This led to approximately 25,000 observations of about 1500 chemists. This preliminary report deals with data on the industrial groups only (approximately 18,000 observations).

Each chemist was observed at two random moments of time each day for 9 consecutive days. Observers were company or university personnel trained and carefully supervised by members of Case's research team. Using “local” personnel for observers considerably reduced the amount of questioning required of the subjects in the observational process. Observations were recorded on the two forms shown in Appendix 1.

Since some refusals by organizations to cooperate was anticipated, a larger sample than was required was selected. Four companies refused to cooperate. This could, of course, introduce some bias in the results if these companies have different characteristics from that of the companies which cooperated. The closest examination we could make for such differences failed to reveal any such characteristics. The reasons for refusal were quite reasonable in most cases and failed to be related to the characteristics of the company as far as we could determine.

So much for the discussion of the plan of the research. The remainder of this paper will deal with the results obtained.

ANALYSIS

The question, “How do chemists actually spend their time?” really has three aspects of equal importance. The first can be stated as: What is the average amount of time spent by chemists in each of the activities classified? This can also be stated in terms of the probability that a randomly observed “chemist-moment” will be classified as a particular activity. This probabilistic form of the question is basic to the mathematical analysis to be described later.

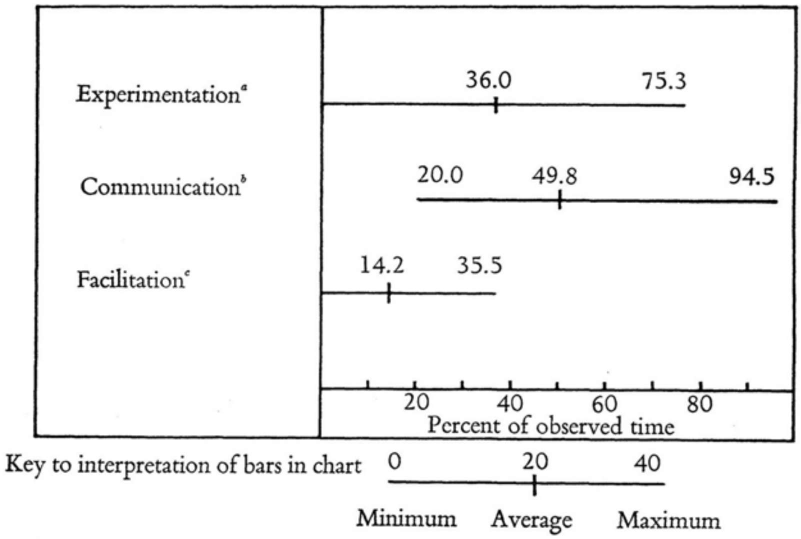
The second aspect of the question as to how chemists spend their time is the consideration of the variability among groups of chemists. This aspect also includes the dynamics of time allocation. Here we are interested in any *relationships* between or among the various classes of activities. This may be stated as: What are the distributions around the averages of time spent, and what functional relationships exist among the averages? One very obvious relationship is that the percentage allocations must add to 100. However, there are other less obvious relationships. This aspect of the time allocation is naturally the most significant in terms of the basic purposes of the study.

These interrelations show where the potential time savings can be obtained. It is not enough to see where the most time is being spent. We must see how the time division among some activities responds to changes in amount spent in other activities. The third aspect of the question concerning the disposition of time by chemists is concerned with the influence of the various environments within which chemists work. Such factors as the type of research in which the work unit is engaged, the availability of scientific literature, salaries, and the professional composition of the work unit may influence time allocations of chemists. More will be said about this point following the discussion of the first two aspects mentioned above.

Environmental data are in the process of being gathered by the questionnaire which appears as Appendix 5 of this report. To date, half of the 50 units included in the study have replied and analysis has just begun. Consequently, there are no preliminary results to report at this time.

TIME ALLOCATION

The results of the 18,000 in area observations show clearly that the conventional picture of the chemist as a man in a white lab smock pouring chemicals into test tubes is a distorted picture. If we lump all experimental, equipment, and data time together we can account for only 36 percent of the chemist's time, while necessary communication (scientific and business) takes up 44 percent of the chemist's work day. Figure 1 shows the division of time among communication, experimentation, and facilitation. The figure also shows the maximum and minimum values for the 100 groups of chemists studied. Most of the following analyses are based on these 100 units which were derived as follows. Each of the 50 groups of chemists was observed 18 times during 9 consecutive days. These 18 observations were divided into "early" (the first 9 observations) and "late" (the last 9) for each group, yielding the 100 units for further analysis.



[a] Includes Equipment setup, Equipment use, and Data treatment.
[b] Includes Scientific communication, Business communication, and Thinking or planning alone.
[c] Includes Personal and social and Miscellaneous.

FIGURE 1. Chemists' time allocation summary.

Figure 2 shows the time division for the eight categories used on form 957–1: Scientific communication; Business communication; Thinking or planning alone; Equipment setup or calibration; Equipment use; Data treatment; Personal and social; and Miscellaneous. The maxima and minima are also shown. It is obvious that Scientific communication is the largest single category. Since our observers were, in most cases, professional chemists and working members of the group being observed, there is little likelihood they would mistake casual conversation (which might contain technical jargon) for Scientific communication. Also, since each observation that was classified as Scientific communication required filling in an additional form (957–2), there was no reason for the observers to err in the direction of overreporting this category. We feel, therefore, that chemists spend *at least* the reported 33 percent of their time in Scientific communication. Also it is the only category that has a minimum other than zero. The data in Figure 2 are presented in Table 1. Except for the two largest categories, Scientific and Business communication, the maxima are approximately four times the average. With the larger categories this drops to 2 to 3 times the average. This, of course, is due to the limiting effect of an upper bound—100%. Removing the effect of this constraint was one difficulty in the development of the detailed analysis.

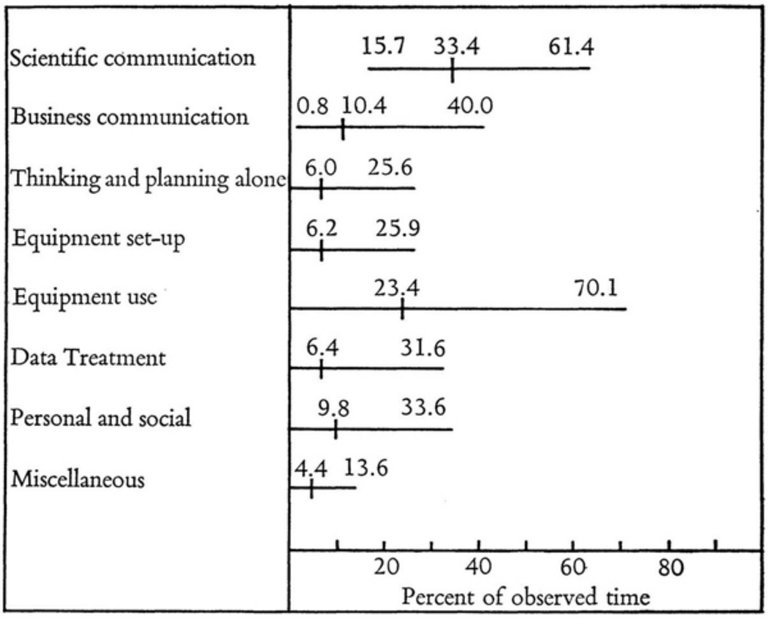


FIGURE 2. Chemists' time allocation, detail.

TABLE 1 Percent time allocation

Activity	*Minimum*	*Average*	*Maximum*
Scientific communication	15.7	33.4	61.4
Business communication	0.8	10.4	40.0
Thinking or planning alone	0.0	6.0	25.6
Equipment setup	0.0	6.2	25.9
Equipment use	0.0	23.4	70.1
Data treatment	0.0	6.4	31.6
Personal and social	0.0	9.8	33.6
Miscellaneous	0.0	4.4	13.6

A further set of time averages was obtained for various aspects of scientific communication. These are shown on Figs. 3, 4, and 5, and in Table 2. The exact definitions of these categories, and all the others used in this analyses are contained in Appendix 2. Some of these relationships are suggestive of experiments, and some lend themselves directly to interpretation. For example, if we construct a fourfold table from Sending oral, Receiving oral, Sending written; Receiving written (Table 3), three things become obvious.

TABLE 2 Percent scientific communication time allocation

Activity	*Minimum*	*Average*	*Maximum*
Total scientific communication	15.7	33.4	61.4
General discussion	0.0	10.3	35.3
Oral, non-discussion	0.0	9.2	28.0
Total			
Written	3.9	14.3	45.0
Unpublished written	0.0	9.5	40.0
Published written	0.0	4.9	18.4
Sending, oral	0.0	4.5	17.7
Receiving, oral	0.0	3.8	19.4
Sending, written	0.0	5.0	15.0
Receiving, written	0.0	7.2	18.4
Retransmittal	0.0	2.7	20.6
Reading articles	0.0	2.6	13.7
Reading for use	0.0	3.9	14.3
Reading for general information	0.0	3.2	18.4
Communication with other scientist, non-chemist	0.0	2.7	16.3
Communication with other company personnel, secretaries, technicians	0.0	7.1	25.7
Communication with chemists	5.3	21.4	54.5

TABLE 3 Sending-receiving vs. written-oral, percent

	Sending	*Receiving*	*Total*
Written	5.0	7.2	12.2
Oral	4.5	3.8	8.3
Total	9.5	11.0	20.5

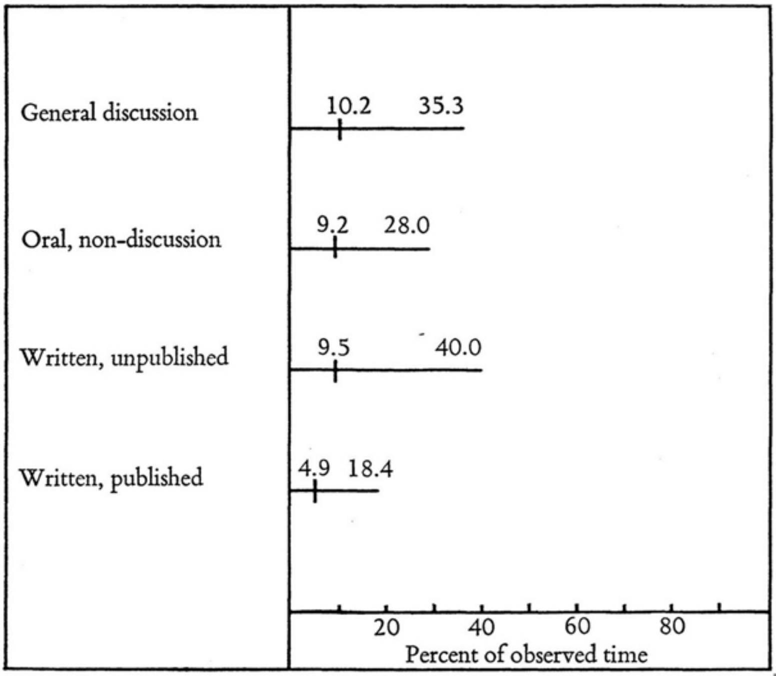


FIGURE 3. Chemists' communication time, allocation to channel.

First, it is apparent that of the four categories here discussed, more time is devoted to receiving recorded communications than to any of the other three categories. This, of course, is to be expected. Not as expected is the fact that while more time is spent in receiving than sending written material less time is spent in receiving oral than in sending oral. Since reading is at least as fast as writing, and since chemists read 1 1/2 times as often as they write (7.2 vs. 5.0) there must be several readers for each item written or else chemists read material written by non-chemists. With respect to oral communication this is reversed. More time is spent talking than listening. Since at least one listener is required for each talker, this must mean that chemists talk to non-chemists more often than they listen to them. Further analysis will explore this in more detail when we deal with the classification of the other people with whom the chemists communicate. This discussion is presented merely to indicate the kind of interpretations to which the data lead. The third, and perhaps least expected result shown by this table, is that written information exchange is used only one and a half times as much as oral. In this table, oral information exchange specifically excludes discussion, and is restricted to actual transfer of information, rather than the give and take of a conversation. In view of the advantages of written communication (speed, clarity, the recipient's choice of time, place, and speed) it seems that the large proportion of oral requires investigation. However, even this ratio (1 1/2:1) indicates that a given message now being transmitted orally would require less than half the total manhours, if it were to be transmitted in writing.

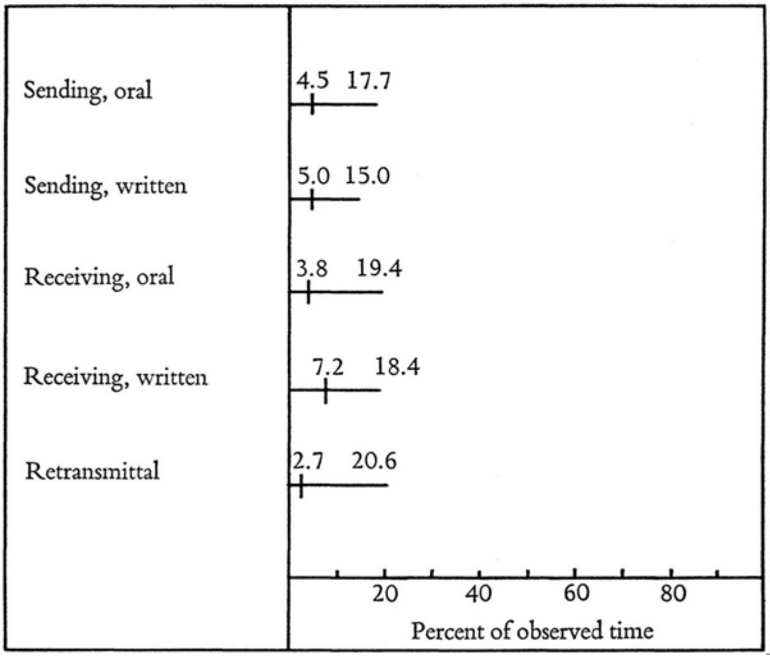


FIGURE 4. Chemists' allocation of communication time by direction of information flow.

INTERRELATIONSHIPS

To investigate the interrelationships among the various activities with respect to the time allocations, it was necessary to allow for the constraint imposed by the fact that the sum of the allocated time must equal 100%. Obviously as more time is spent in, say, scientific communication, less is left to be divided among the remaining activities. The definition of independence among activities used is: Activity *Y* is independent of activity *X,* if and only if, as *X* changes from the average, *Y* receives (or loses) time *in proportion* to the average *Y.* Mathematically this means that the expected value of *Y* is $\overline{Y}(1-X)/(1-\overline{X})$, where $\overline{Y}$ =average value of *Y*, and $\overline{X}$ =average value of *X*. It turns out that this is the same value to be expected under the multinomial hypothesis. See Appendix 4 for the mathematical statement of the multinomial and the derivation of the expected value under constraint.

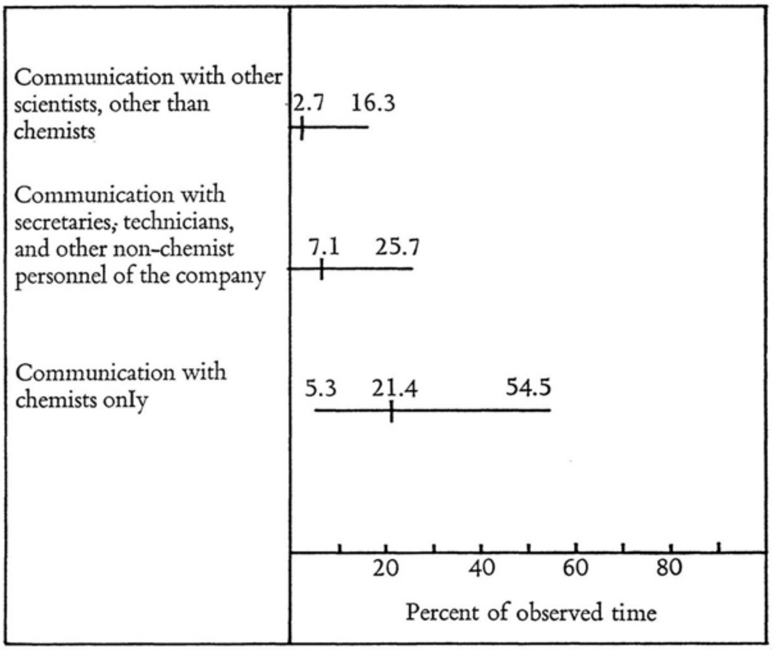


FIGURE 5. Chemists' allocation of communication time by person communicated with.

With this expected value, interrelationships can be explored to see if, in fact, the actual values tend to exceed or to be less than the expected values. As an example, we will consider the hypothesis that the amount of time spent in Equipment use is dependent on the amount of time spent in Scientific communication. Figure 6 shows a plot of the raw data, the percent of time spent by each observed unit in each of these two categories. Here there is a definite indication that as Scientific communication increases, Equipment use decreases. But how much of this is due to the limit of 100% as a total?

Figure 7 shows the data replotted in terms that exclude this effect. The abscissa shows the deviation for each unit from the average Scientific communication (33.4%). The ordinate shows the deviation for each unit from the *expected value* of Equipment use; expected according to the definition of independence. This clearly shows that while Equipment use is dependent on

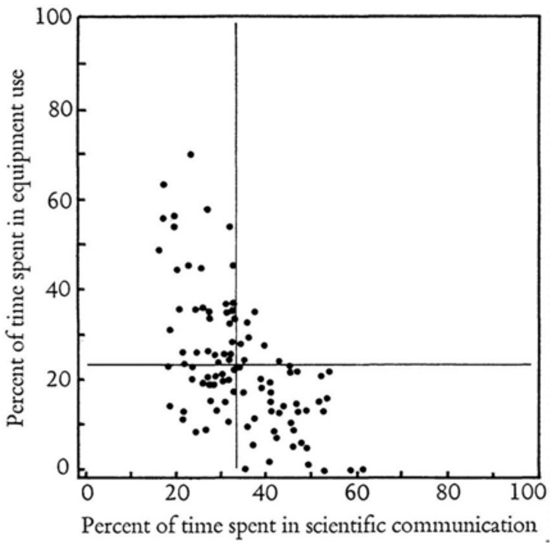


FIGURE 6. Relationship between Equipment use and Communication.

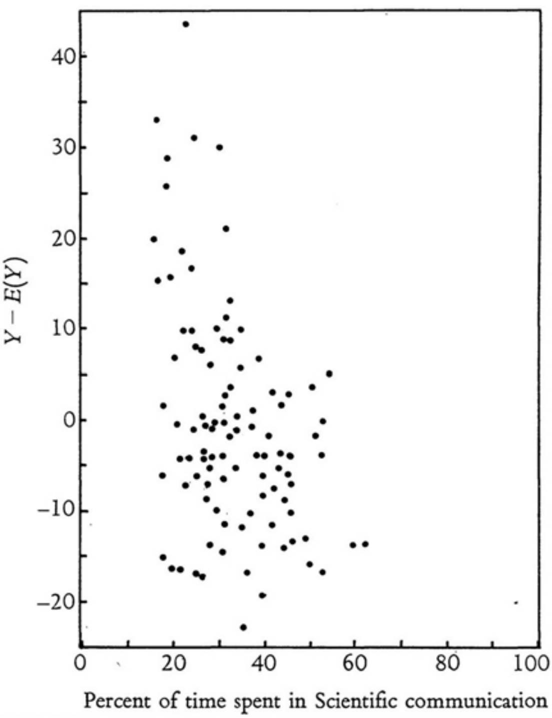


FIGURE 7. Unexplained Equipment use. *Y,* percent of time spent in Equipment use; *E (Y),* predicted percent of time spent in Equipment use. (See Appendix 4.)

Scientific communication, much of the apparent dependence was due to the 100% limit. As more time is spent in scientific communication, *even less than would be expected* is devoted to Equipment use. And as less time is spent in Scientific communication, Equipment use gets more than its share of the residual. Figure 8 shows that there is no effect on amount of Scientific communication by changes in equipment use.

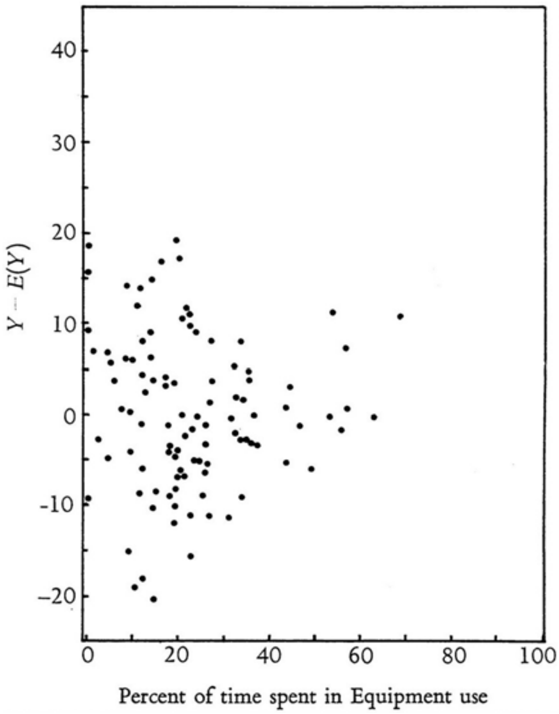


FIGURE 8. Unexplained scientific communication. *Y,* percent of time spent in scientific communication; *E(Y),* predicted percent of time spent in Scientific communication. (See Appendix 4.)

In addition to the charts, a statistical analysis using chi square is shown in Table 4. This is the 2 by 2 cross-contingency table showing whether the number of points in each quadrant of the graph (as in Figs. 7 and 8) could have arisen due to chance, when in fact the control variable exerts no influence on deviations from the expected value of the dependent variable.

An additional series of tests were run to investigate the relation between these variables and some of the characteristics of the units observed. Appendix 5 contains a copy of a questionnaire sent to each cooperating company. At present only enough returns are in to permit partial analysis, but one of the most striking results is that none of the variables of time allocation is dependent on the number of chemists in the observed group. This group size varied from 5 to 111. Table 5 shows the relation between group size and each of the eight gross time allocations. The hypothesis that large units require more internal communication is scarcely borne out by these data. While some of the variables do show a slight dependence on size, it is not so large as was expected.

TABLE 4 Chi-square tests on deviations from expected value of dependent activity

Control Activity \ Dependent Activity	SC	BC	T and PA	ESU	EU	DT	P and S	M
SC		P=.005 + [8.088]	P=.005 + [9.105]	[0.710]	P=.050 — [4.624]	P=.025 + [5.602]	[0.145]	P=.200 + [2.064]
BC	P=.200 + [2.681]		P=.100 + [3.438]	[0.040]	P=.010 — [7.798]	[0.040]	[0.114]	P=.200 + [2.532]
T and PA	P=.005 + [9.246]	[0.944]		[0.174]	P=.200 — [2.356]	[0.451]	[0.008]	[0.272]
ESU	P=.025 — [5.368]	[0.270]	[0.307]		P=.025 + [5.999]	[0.066]	[0.626]	[1.040]
EU	[0.060]	P=.050 — [4.781]	[0.394]	P=.025 + [5.557]		[0.004]	P=.05 + [3.937]	P=.100 — [3.578]
DT	P=.025 + [6.742]	[0.046]	[1.454]	[0.061]	[0.466]		[0.568]	[1.346]
P and S	P=.200 — [1.820]	[0.137]	P=.200 + [1.923]	[0.413]	[0.828]	[0.413]		P=.100 + [3.619]
M	P=.050 + [4.431]	P=.200 + [1.944]	[0.277]	[1.104]	P=.200 — [2.490]	[1.343]	P=1.00 + [3.457]	

SC, Scientific communication; BC, Business communication; T and PA, Thinking or planning alone; ESU, equipment setup and maintenance; EU, Equipment use; DT, Data treatment; P and S, Personal and social, M, Miscellaneous.
In each square of the table: P=0.005, means the probability is 0.005 or less than the degree of dependence shown by the data is due to chance. If the probability was greater than 0.200, it was omitted; + or ` , the relationship between the control and dependent variables was positive or negative; [8.088], the value of chi square for the indicated relationship, with one degree of freedom.

Additional analyses under way will investigate the relationships among the several components of communication, as well as with company characteristics such as basic vs. applied research, number and type of persons communicated with, and library facilities.

TABLE 5 Group size as it affects time allocation

Activity class	*Large[a] companies above average time allocation*	*Small[b] companies above average time allocation*	*Probability[c] of difference at least this large*
Scientific communication	15	9	0.10
Business communication	10	10	1.00
Thinking and planning alone	5	11	0.10
Equipment setup	11	5	0.10
Equipment use	8	9	0.90
Data treatment	10	9	0.90
Personal and social	11	10	0.90
Miscellaneous	16	8	0.025
Out of area	15	10	0.20

[a] Companies with more than 20 "observed" chemists.
[b] Companies with less than 21 "observed" chemists. Total number of companies is 50.
[c] Computed from chi square 2×2 tables.

The major purpose of this phase of the analysis is to determine what will be the effect of changing the amount of communication. If the relative value of the various activities can be approximated (perhaps through experiments) then communication time can be increased at the expense of less valuable activities, or decreased in favor of more valuable ones. In any event, the value of communication (or any part of communication) must be meaningfully related to only those activities that vary non-randomly with it. When completed, a direction will have been established for the development of a measure of the value or worth of various kinds of written and oral communication.

Naturally, since the analysis is still in progress, the conclusions and interpretations are tentative; they will be confirmed, rejected, or modified on the basis of the further results, which will, we trust, be available in the very near future.

INTRACOMPANY RELATIONSHIPS

Since there were 18 successive sets of observations made on each company, it is feasible to investigate the intracompany variability. One examination was made to see if there was any trend either in the averages or in the dispersion. This analysis was conducted on a sample of ten of the fifty companies, and a modified form will be extended to the rest. The results indicate that there was no time trend. This is evidence that our training sessions for the interviewers was adequate to bring them to an acceptable level where no further learning occurred during the data collection.

A second analysis of the intracompany data was performed to test the hypothesis that the 18 sets of data we obtained were, in effect, 18 random samples from the same population.

In order to run the test, the form 1 frequency data for each unit (see Appendix 1) were grouped into 6 activities within each of the 18 rounds: scientific communication, business communication, thinking or planning alone, equipment setup and data treatment,[1] equipment use, personal and social. The other form 1 classifications, "none of these" and "out of area," were not included in the test.

By classifying the data by activity and round number, a 6 by 18 matrix was formed. The row and column totals were used to compute expected values within each cell of the table. X^2 was computed with 5×17=85 degrees of freedom.

Possible interpretation of the results are:

1. Low values of X^2 may be explained, as the result of task specialization. One chemist works mainly on equipment, another searches the literature, etc. This violates the "equal probability" assumption on which the test is based. It should be noted here that the hypothesis under test is whether all chemists *within a unit* are alike. This makes no statements about differences among companies.
2. High values of X^2 may be explained as the result of chemists performing certain activities in groups. Discussion, end of the month reports, and "personal and social" are obvious examples of this possibility. Here the assumption of independence (and hence the "random sample" assumption) is violated.
3. Certain of the smaller units have low frequencies for some of their activities. The chi-square approximation often fails under such conditions and spurious values of X^2 (usually too high) may result.

With X^2 for 10 units calculated, all the values have been high, the *lowest* being 87.3. The results are tabulated in Table 6.

The effect of low frequencies is important with some of the smaller units,

TABLE 6 *Chi-square results of intracompany analysis*

Size of unit	*Number of observations*	*Value of* X^2	*Probability* $X^2 \geq \nu$
9	147	104.40	<0.10
20	280	87.32	≈0.40
25	313	162.20	<<0.0001
31	429	98.61	<0.25
34	461	112.46	<0.025
35	508	118.13	=0.01
40	623	115.30	<0.025
63	814	135.14	<0.0001
99	1183	95.67	<0.25
111	1433	163.05	<<0.0001

[1] These two were grouped together because of their similarity of function and to increase the expected frequencies within each classification.

but the most important factor seems to be grouping. This is evidenced by the fact that large units show no tendency to have a lower X^2 than small (low frequency) units, and also by the consistently high contribution of “personal and social” to X^2, even when the expected values exceed 4 in every box.

PLANS FOR ADDITIONAL ANALYSES

In summary, two preliminary analyses have been conducted. The first was a search for interrelationships between classes of activity. In this analysis all observations for each group of chemists were combined into single estimates of the proportions of time spent in each activity. Then each group was treated as a single data point in looking for relationships between activities. The assumption was that, if there were no relationship between two activities, then, as the proportion of time spent in one of these activities moves away from the average proportion over all groups, the proportion of the *residual* time represented by the second activity will remain constant and equal to the average proportion of the residual for that activity over all the groups. As yet, the analysis has not isolated any strong relationships which would deny this assumption.

The second analysis tested the assumption that the chemists within a group can be viewed as a single population, each member of which distributes his time in accordance with the allocation typical of the group as a whole. If this assumption were true, the activity pattern during each of the 18 observation periods within a given group should be randomly distributed about the over-all activity pattern for the group. This assumption was not supported by the test. Instead there appeared to be a substantial “grouping” effect, that is, chemists seem to work at tasks in groups so that what one chemist does influences what another is doing.

A third analysis which is just starting will investigate whether there are significant differences in time distributions between groups. If there are, these results taken together with those from the second analysis would indicate that the environment within which a group of chemists work has a strong influence on how the group distributes its time. The environmental data being gathered should be useful in isolating those factors which affect the allocation and should indicate the nature of the relationship between a given factor and the group's activity pattern. Knowledge of these factors and their influences on time allocation may provide means to manipulate the allocation in directions favorable to increased productivity.

A number of other potentially important analyses are contemplated. First, although a direct attempt to construct a measure of effectiveness based on scientific productivity has been purposefully avoided, there are a number of broad-gauge *effects* of productivity which can be utilized in a search for useful hypotheses. For example, it seems reasonable that a chemical company which has experienced a significantly more rapid increase in sales volume over the past decade than could be explained by the normal expansion of the economy could very well have done so as a result of introducing successful new products. If these products were developed by the company's research staff, then it must be considered relatively high on the scale of scientific productivity, in the applied sense, at least. Assuming a relatively stable protocol for scientific behavior within companies, comparisons between the activity distributions of chemist groups in highly productive companies with those in the other end of the scale may bring to light important effects. Further, adjusting for differences in work group environment may make these differences more apparent.

A second major analytical segment will be concerned with evaluating the relative efficiencies of various means of communication, and the effects of using more or less of the efficient media on the areas of activity from which scientific productivity must come. As an example of this type of analysis suppose that, in the area of scientific communication, the total amount of information received through written media is higher per unit of time spent in reading and writing than is the amount of information received orally per unit of time spent in talking and listening. Now suppose that the groups of chemists are arranged in descending order according to the proportion of each group's total scientific communication time that is concerned with written communication. One possible result might be that these groups which spend more of their communication time in using written communication also spend a higher portion of non-scientific communication time in a combination of Thinking and planning alone, Equipment use, and Data analysis than do those whose Scientific communication is predominantly oral. If so, extension of this analysis into the content of the written communication should provide additional understanding of the relationship and, consequently, a higher probability of being able to modify the relationship favorably through changes in content.

In addition to using the content of communications in conjunction with the relative efficiency analysis described above, content analysis per se may provide broader comprehension of the scientific activity process. Relationships such as that between the portion of time spent in reading chemical abstracts versus the portion of time spent reading professional books and articles may indicate whether abstracts are used as a substitute for the professional literature on which they are based or whether they are used as guides to this literature. In other words, insight may be gained into the function of one form of scientific communication with respect to the other forms. Furthermore, relationships can be sought between the form and content of a communication and the other aspects of the group's activity pattern.

ACKNOWLEDGMENT

This project involved the efforts of many including Beverly Bond, Richard E. Deal, Philip Forsyth, Robert J.Frame, Michael Leyzorek, Joseph F.McCloskey, B.H.P.Rivett, George W.Summers, and Ronald W.Wolff, all from Case. The great contribution of Dr. Alberto Thompson and Mrs. Helen Brownson of the Office of Scientific Information, National Science Foundation, should also be cited. The study would not have been possible without the cooperation of the American Chemical Society and the many companies and universities and their personnel who permitted themselves to be observed.

APPENDIX 1

No: Round: Observer:

Scheduled start time:

Actual start time:

ACTIVITY	NUMBER OF OBSERVATIONS	TOTAL	
Scientific communication			11 12
Non-scientific business communication			13 14
Thinking or planning alone			15 16
Equipment set-up and maintenance			17 18
Equipment use			19 20
Data treatment			21 22
Personal and social			23 24
None of these			25 26
Out of area			27 28
Total			29 30

Casor form 957-1

No: Round:

DESCRIPTION OF SCIENTIFIC COMMUNICATION

Hearing question	from	1
Reading question	from	2
Reading for use	from	3
Reading for general information	from	4
Hearing information	from	5
Working out material	from	6
Editing material received	from	7
Writing information	to	1
Telling information	to	2
Writing question	to	3
Asking question	to	4
General discussion	with	5
Discussion about a received communication	from with	6
Reading for retransmittal	from to	7
None of these (write in)		8

NUMBER AND TYPE OF PEOPLE INVOLVED IN THE SCIENTIFIC COMMUNICATION

From		To, With
☐	Mathematicians, statisticians	☐
☐	Physical scientists	☐
☐	Chemists, same unit	☐
☐	Chemists, different unit	☐
☐	Biologists, medical men	☐
☐	Behavioral scientists	☐
☐	Secretaries, technicians, etc.	☐
☐	Other company personnel	☐
☐	None of these (write in)	☐

FORM OF THE SCIENTIFIC COMMUNICATION

From		To, With
1	Oral	7
2	Unpublished written	8
3	Book	9
4	Article	0
5	Abstract or review	X
6	None of these (write in)	Y

Casor form 957-2

APPENDIX 2 DATA COLLECTION INSTRUCTIONS FOR PROJECT 957

1. *List of materials*

a. 25 copies of Casor form 957–1 and envelopes,

b. A supply of Casor forms 957–2.

c. A set of instructions.

2. *Observation procedure*

a. An observation is made on each selected chemist and consists of looking at what he is doing when you first come into his office, lab. or meet him in the hall on either your morning or afternoon round. Each observation is checked on form 957–1 and if it is “scientific communication,” then a form 957–2 is filled out.

b. The list of “chemists” for this study is made up of full time employees who:

(1) are members (not student affiliates or student members) of A.C.S., or

(2) have a B.S. or higher degree in Chem. or Chem. E., and are currently working in chemistry (utilizing their technical training), or

(3) have a bachelor's degree in some other science (math., biol., statistics, etc.), and have been working as chemists for at least one year, or

(4) have no degree in science, but have been working as chemists for three years.

This definition is not intended to include technicians, machinists, laboratory assistants, etc. “Working as a chemist,” means that the level of skill and training required is that usually indicated by a B.S. degree.

c. The chemists on the list are those who have their desks (or can usually be found) in the area that contains the traced chemist. This area is chosen to be feasible for observation, and contains no less than five chemists, but may contain as many as 100.

d. An observational round starts at the time shown as “Scheduled start time” on the top of form 957–1. You go around the area and observe the chemists' activities. It is easier if you carry a list of the chemists and cross off each one as you see him. If there is a chemist you don't see, find out where he is, and if it is feasible, go and observe him, (he may be down the hall at another lab.). If he is out of the building or cannot be located, check “Out of area” on form 957–1.

e. After each round, check the forms to see if you have filled out all the columns, and that you have a form 957–2 for each check under “scientific communication” on form 957–1, put the completed forms in an envelope, and relax (or go back to your regular work).

3. *Casor form 957–1. (See sample.)*

a. There is one form 957–1 for each round. There are two rounds a day, one in the morning and one in the afternoon. This form will contain a mark for each chemist on the list. Either he will be classified in one of the first eight rows (Scientific communication to “None of these”) or will be, “Out of area.”

The "total" number of checks must always be equal and be the number of chemists on your list. Every chemist must be accounted for.

b. The heading of this form contains five items (see sample). The first, "No. _____" will be filled out when you get the forms. It indicates your organization's code designation. The second item, "Round _____," is to indicate which set of observations are recorded on the form. This is a two digit number. The second digit will always be "1" or "2." "1" shows that the round was a morning round, "2" indicates an afternoon round. The first digit indicates the day. The first day on which observations are to be made will be day 1, the second day will be "2," etc. Thus round "3 2" will be the afternoon round on the third day of data collection; round "8 1" will be the morning round on the eighth day, etc. "1 1" will be the first round and "9 2" will be the last.

c. In the item "Observer _____" put your initials (or the initials of the person who makes a round in your place). This is so that if there are any questions they can be referred to the correct person.

d. The next item is "Scheduled start time _____." This is filled out and is either on the hour or the half-hour. These times were picked to give a representative picture of your working day, allowing for lunch. In the next blank, "Actual start time _____," enter the actual time you start the round, to the nearest minute. We fully realize that you have your regular duties to attend to, and that there will be some rounds that will be started early or late. Please enter the *actual* time you start the round, even if it is quite different from the scheduled time. If you cannot make the round in the *same half day* as it is scheduled, write "Missed" in the "Actual start time _____" blank, and a make-up round will be added at the end of the regular nine-day period. e. The body of the form consists of ten rows, a space for counting marks for each row, and a "total" column. Do not use the column at the extreme right of the form (the one that starts "11, 12, 13"). In the large space, using single strokes with a diagonal for five to record the individual observations. For example, "///" is three, "𝍸 //" is seven, and "𝍸 𝍸 //" is 12, etc.

f. In the total column, record the number of marks in each row, and enter this number under "total." This will make adding the grand total easier, and will facilitate further analyses.

g. The categories can best be defined by examples.

(1) "Scientific communication" is meant to include all talking, listening, reading, writing, and discussion concerning technical or scientific matters related either to the job or to outside professional activities.

Examples are: reading an article, handbook, memorandum, etc.; discussing work, project status, etc., talking on the phone to someone about a technical matter; writing an article, book, report, memorandum, etc.; editing or reviewing material for publication.

(2) "Non-scientific business communication" consists of all the necessary communication (business and professional) that is not of a scientific or technical nature. This category includes supervisory and administrative communications, letters, memos, and discussions relating to conducting the business activities or outside professional activities.

Examples are: discussing personnel, office space, price, delivery, vacation schedules, and staff meetings.

(3) "Thinking or planning alone" is the category for activities such as writing your own notes, working out a plan on paper, (or blackboard), or just sitting and thinking (about scientific matters). Some chemists work best in groups, some best alone. This category is not to be confused with "scientific communication," which may well be used for activities done alone. If there is communication with someone else, by writing, reading, or phone, even though the chemist is alone, it is "scientific communication." If he is alone and thinking or planning (not communicating with anyone else) then this category (3) applies.

Examples are: designing an experiment, working out an idea, sketching an apparatus layout *for his own use,* and thinking thru a problem.

(4) "Equipment setup and maintenance" and

(5) "Equipment use" are two very important categories and will account for most of the observations. They include working with chemical and scientific apparatus. The two classifications distinguish between work that is preliminary (including calibration) and that which yields actual data. The setup and maintenance are necessary, but are to be coded separately from the use of the equipment.

Examples are: calibrating a *p*H meter, bending tubing, recording readings on a data sheet, washing flasks, and titrating a solution.

(6) "Data treatment" is to be used for the analysis of numerical data such as calculating formulae, drawing graphs, making up tables, and charts. This category is for the statistical, numerical, mathematical treatment of observations, not for mere transcription or recording (see examples in 4, 5).

Examples are: adding numbers, calculating averages, and drawing graphs.

(7) "Personal and social" is for the ordinary daily activities that chemists perform that are not actually part of their jobs. They do them because they are humans as well as chemists. This includes coffee breaks, talking about the weather, TV, politics, or the wife and children. Also included are personal or social communications such as writing personal letters, checks, calling friends or family. Be sure there is no attempt either to minimize or to exaggerate this category. There is not the slightest implication that personal and social activities are undesirable. They are part of the everyday business of living and getting along with others.

(8) "None of these" is a miscellaneous category for cases that do not fall into any of the others. It includes walking in the halls, waiting, and straightening up a desk. In every case where it is used, make a note on the back of the form explaining what the observation was. Use this category only if you cannot fit the observation into any of the other-categories.

(9) "Out of area" is for those cases where you cannot locate the chemist or you know that he is out of the building, home, sick, on vacation, on a business trip, etc. If you can categorize his activity then do not list him here, but put the check in the appropriate box. If he left the office for a doctor's appointment, for example, check "Personal and social." If he

went to the library check “Scientific communication,” but on the 957–2 just write in “At Library, not observed.” You may have to ask to find out where he went.

(10) For “None of these” write details on back of form.

4. *Casor form 957–2.*

One copy of this form must be filled out for each mark under “Scientific Communication” on form 957–1.

a. The heading “No: _____, Round: _____” should be filled out to agree with the same entries on form 957–1. These forms will be kept together at the end of each round. There will be one 957–1, and as many 957–2's as there are checks under “Scientific communication” on 957–1.

When filling out a 957–2, there must be one entry in the first column, and two entries in the second column, one under “Number and type of people involved in the scientific communication” and one under, “Form of the scientific communication.”

b. In the first column an entry is made by *circling* the number at the right of the entry, after the word “from” or “to” or “with.” The purpose of the word is to remind you that the entries in the second column go on the left or right, depending on whether they refer to a “from” or a “to, with.” In this second column, an entry will be a *number* written in by you in the appropriate box. This number tells how many people the observed chemist is communicating with. If he is discussing something with a group, then you write in the number of people (do not include the observed chemist in this count). On every form 957–2 there will be at least one entry in this section. There may be several, since the other people being communicated with may be distributed among the categories; for example, some biologists and some chemists might be discussing something with the observed chemist. If the observed chemist is reading, then the number and classification of “Number and type of people …” would refer to the author or authors. It is extremely important that these numbers be accurate. The only time numbers will not be used is when material is being sent to a large and undetermined audience, as in the case of an article being written for a journal or a speech being prepared for a professional meeting. In that case, use the letter *Y* and put it in the appropriate box or boxes. If you have no idea as to the composition of the audience, use “none of these” and describe the situation on the back of the form.

In the last section (bottom) of column two, an entry is made by circling a number (or *X* or *Y*).

In the two places (column one and bottom of column two) where there are numbers printed in, an entry consists of circling the number. In the other case (top half of column two) where there is no number, write one in.

c. Column 1. Description of scientific communication

(1) *Hearing question* refers to occasions where the observed chemist is listening to an oral request or question. The actual words do not have to be in question form. It may be a suggestion, or request. The important aspect for this classification is that the answer or reply would normally contain scientific information. This category is *not* to be used for casual questions

that occur in discussion like, "Don't you think so, Joe?" or, "Well, then, is it agreed that well get together at 10:00 on Tuesday?"

This classification should be used for oral requests for technical information, for example: "Joe, would you know what journal had that article by Klumpenmeyer on semi-captive radicals?" This category is also used for questions about apparatus, calibration, etc. If the question comes up in a *discussion, do not use this category,* use one of the two discussion categories. This category is for questions or requests for information that are the purpose of whole communication. Note that the entries in column 2 will be in the "from" column and that "(1) oral" will be recorded at the bottom of column 2.

(2) *Reading question* is the same as hearing question, except that it refers to reading a question or a request for information.

(3) *Reading for use.* This classification is for the reading of memoranda, articles, books, etc. when the purpose of the reading is to find some specific information of direct use on a current task. Examples include looking up values in a handbook and checking a formula in a text or reference book.

(4) *Reading for general information.* This category is for all the scientific reading that is done for background, for its *general* value, because you get the journal anyway, or just because it is interesting. The reading must be scientific or a 957–2 would not be filled out at all. (*Popular Mechanics* is not scientific literature.)

(5) *Hearing information* is used for two kinds of situations. One is in the answer or reply to a question. Remember this refers only to those situations where the main purpose of the interchange was the answering of a particular question or request, not those questions arising during a discussion. The second situation is in listening to a lecture, briefing session or talk where there is little chance to break in with questions or remarks. This use of "hearing information" is used to denote a one way communication system. If the exchange is two way, then the category to be used is "discussion."

Examples of "hearing information" are: Listening to the answer to a technical question, being briefed on a new technical development by a visiting expert, and listening to a report of a professional meeting by someone who attended.

(6) *Working out material* is used for those situations where the observed chemist is trying to understand something he has read or heard. He may be using a blackboard or pad to see if a certain reaction will follow as Hemplemeyer says in his article. In the later parts of the recording form the data refer to the source of the material he is working out.

(7) *Editing material received* is for those cases in which the observed chemist is proofreading, editing, refereeing, or otherwise working with material for publication (either for journals or books, or for publication within the company). This may be something he has written himself or something that someone else wrote, but it is now being readied in final form,

not still being written. In the last section of 957–2, use the category that refers to the ultimate form. If a chemist is editing material to be submitted as an article to A.C.S., this would be called "article" even though it is as yet unpublished. If he is reworking a speech for future delivery it is "oral" even though the speech may eventually be published.

All the above categories are "from" and must have entries under "From" in the second column on the form.

(8) *Writing information* is used for all written scientific communication including letters, memoranda, reports, articles or books. It refers to outgoing information only and not to written requests for information.

(9) *Telling information* is the same as "writing information" except it refers to oral methods rather than written. If one person is "hearing information" (as discussed above) then someone else must be "telling information."

(10) *Writing question* refers to writing a request for information.

(11) *Asking question* is the same as "writing question" but is oral, not written. It is the other half of "hearing question."

The previous four categories are all "to," and must have their entries in column two on the right-hand side, "To, With."

(12) *General discussion* is that part of scientific communication that goes on in groups. It is not "one way" (like hearing or telling information). The discussion may be about some highly specific point, but it must be a real discussion, not a lecture. Note that this category is "with," and requires the number and kind of co-discussants.

(13) *Discussion about a received communication* is similar to "working out material" but refers to a "bull session" where the participants are trying to understand or put to use something they were told or read. This category is "from" *and* "with." It requires entries on *both* sides of column two.

(14) *Reading for retransmittal* refers to the case of an observed chemist reading an article or book, etc., to report on it (either orally or in writing). In this case as in the previous one, there will be entries on *both* sides of column two, "From" and "To."

(15) *None of these* is for any difficult case that doesn't fit in. Be sure to write in the details on the back of the form.

d. Column 2. Number and type of people involved in the scientific communication.

In column two of form 957–2, an entry is a *number*. Write in a number in the box, using either the left or right set according to the "from" or "to" in the entry you circled in column one.

(1) *Mathematicians, statisticians* include all the mathematical sciences, experimental designers, and geometers.

(2) *Physical scientists* include physicists, metallurgists, astronomers, optical scientists, meteorologists, and electronics men.

(3) *Chemists, same unit* refers to chemists who are in the same project group, or in the same administrative unit (where this is defined as narrowly as possible). The intent here is to include men working on the same project, or a group of closely related projects.

(4) *Chemists, different unit* refers to other men who are chemists, but who are working on different projects or different types of problems. If a biochemist working on new insecticides is talking with a physical chemist working on artificial crystals, they are in different units. (For “unit” you can read “branch of chemistry.”)

(5) *Biologists, medical men* include what are often called the life scientists. These are biologists, pharmacologists, zoologists, botanists, medical men, etc.

(6) *Behavioral scientists* are the group containing psychologists, economists, political scientists, semanticists, etc.

If you are in doubt, write in the name of the science and then the number of such scientists involved.

(7) *Secretaries, technicians, etc.,* include students, lab. assistants, and machinists. These are the people that directly help the chemist get his work done.

(8) *Other company personnel* are your company's sales men, executives, administrators, staff assistants, etc.

(9) *None of these* includes customers, visitors, outsiders of all kinds. Be sure to write in an explanation on the back of the form,

e. Form of the scientific communication.

The last section of form 957–2 deals with the nature of the communication. Again check in the appropriate side (“From” or “To, With”). In the case of “Reading for Retransmittal” there will be a check on *each* side, one for the nature of the reading, one for the form of the retransmittal. “Unpublished written” includes letters, memoranda, and reports for company or client use. “Published” (book, article, abstract or review) means that the material will be generally available to chemists (often with difficulty—as a Ph.D. thesis).

In every case where anything but “oral,” or, “Unpublished written” is checked a reference must be put on the back of the form showing the title, journal, issue, or date, and page number. This is extremely important. Get the complete reference, so if we go to a library we can find it.

In every case on the form where “None of these” is checked, write in the details on the back of the form.

5. *Security*

Since no names or code numbers are associated with individual chemists, there is no possibility (and certainly no intention) of identifying any chemist.

A copy of the final report will be given to your organization, but it will not report data for any individual or company. No one in your organization except you will see the recording forms. Thus the integrity and individuality of your chemists is guaranteed.

APPENDIX 3

I. The Universe to be sampled (target universe) is the set of Chemist-Minutes constrained by the following restrictions:

1. Continental U.S.
2. 5-day week, normal work hours/day.
3. Work in one of the 150 metropolitan areas (1950 census definition) listed in Table 73, Detailed Characteristics, U.S. Census of Population.
4. Work for a unit (smallest administrative unit in a company) employing 5 or more chemists (members of A.C.S. as of June 1, 1957). [*Note.* Restriction 3 limits the sampled universe to about 80% of the target universe. Restriction 4 limits the sampled universe to about 85% of the target. If they are uncorrelated, then, the restriction is to about 68%. However, there is probably some positive correlation (larger companies are more likely to be in metropolitan areas) so 68% is a minimum estimate.]
5. The time period—July to October 1957.

II. The procedure used to generate a sample of Chemist-Minutes is as follows:

1. The hundred fifty metro areas were ordered from most chemists to least chemists. A systematic stratified sample of 10 areas was drawn which yielded (1) New York, (2) New York, (3) Chicago, (4) Philadelphia, (5) San Francisco, (6) Wilmington, (7) Buffalo, (8) Albany, (9) Kalamazoo, (10) Trenton. This sample was based on the 1950 census data (Table 73, Detailed Characteristics).
2. In each metro area, the names of 50 chemists were drawn at random (systematic random sample—no stratification). Data used was the current A. C. S. roster.
3. Each name drawn was checked to get number of chemists employed by the firm (same general address—to treat individually, multi-plant firms) and occupational classification of chemist. If firm employed ≥5 and occupation was in (see Appendix A for list of occupational classification), then firm (unit) was included in the sample. This operation was continued until approximately 100 chemists (totaling the number of "in" chemists employed by each unit) were obtained.
4. This generated a sample of approximately 60 firms and 100 chemists.
5. Each chemist in each unit was observed at two random times (one before lunch, one after) for nine consecutive work days. These observations (approximately 18,000) form the basis of the further analysis.

APPENDIX 3A

The sampling procedure yielded a sample with the following characteristics: number of companies, 42; number of observed units, 50; number of chemists, 1305.

Breakdown by Metropolitan area:

Metro	*Companies*	*Units*	*Chemists*
New York	11	15	257
Chicago	6	6	159
Philadelphia	2	2	114
San Francisco	6	6	147
Wilmington	2	2	160
Buffalo	4	4	89
Albany	6	7	179
Trenton	4	5	96
Kalamazoo	1	3	104

The sample was designed to obtain approximately 100 chemists in each area (200 in New York since it was considered two areas). The size of the units varied from 5–111 chemists with most of them lying within a range of 10–40.

APPENDIX 4 THE USE OF THE MULTINOMIAL ASSUMPTION

Since we have a universe of chemists one assumption we may make about the underlying structure is that each and every chemist has the same probability for any particular activity as any other chemist, and that these probabilities do not change with time.

Then, if we were to observe a group of chemists, we could predict, on the average, the kinds of observations we should get.

Thus, if there are k different, mutually exclusive and exhaustive activities, and associated with each one is a probability $(P_1, P_2, P_3, \cdots, P_i, \cdots, P_k)$ such that

$$\sum_{i=1}^{k} P_i = 1$$

then if we observe n chemists at random, the probability that we will get exactly x_1 observations of the first activity, x_2 of the second, etc., and x_k of the kth activity is given by:

$$\Pr\{x_1, x_2, x_3, \cdots, x_i, \cdots, x_k\} = \frac{n!}{x_1!\, x_2!\, x_3! \cdots x_k!} P_1^{x_1} P_2^{x_2} \cdots P_k^{x_k}, \qquad \sum_{1}^{k} x_i = n$$

This is the multinomial distribution.

Our concern in this situation, however, is to determine the distribution of x_1 given x_2, i.e., $\Pr\{x_1|x_2\}$. This can be done simply by applying conditional probability to the multinomial for x_1, x_2, and x_3.

Letting $M(x_1, x_2)$=the multinomial for x_1, x_2, we have

$$\Pr\{x_1 \mid x_2\} = \frac{M(x_1, x_2)}{M(x_i)}$$

which equals

$$\frac{n!\, P_1^{x_1} (1 - P_2)^{n-x_1-x_2}\, x_2!\, (n - x_2)!}{x_1!\, x_2!\, (n - x_1 - x_2)!\, n!\, P_2^{x_2} (1 - P_2)^{n-x_2}}$$

which reduces to the expected value of $x_1|x_2$:

$$E(x_1 \mid x_2) = \sum_{x_1=0}^{n-x_2} \frac{x_1\, (n - x_2)!\, P_1^{x_1}\, (1 - P_1 - P_2)^{n-x_1-x_2}}{x_1!\, (n - x_1 - x_2)!\, (1 - P_2)^{n-x_2}}$$

This can also be expressed as

$$\frac{(n - x_2)\, P_1}{1 - P_2}$$

or in percent terms

$$P_1 \frac{1 - x_2}{1 - P_2}$$

which is the computational form used in the body of the report.

The variance of

$$E(x_1 | x_2) = \frac{(n - x_2)\, P_1\, (1 - P_1 - P_2)}{(1 - P_2)^2}$$

or, the standard deviation, in percent terms is:

$$\left(\frac{1 - x_2}{n}\right)^{1/2} \left(\frac{P_1\, (1 - P_1 - P_2)}{(1 - P_2)^2}\right)^{1/2}$$

These two forms, the percent forms for expected value and standard deviation, are the ones utilized for the computations in the report.

APPENDIX 5 SCIENTIFIC COMMUNICATION STUDY—QUESTIONNAIRE

1. Would you please classify the *major* activity of the “observed chemists” as:
   - ☐ basic research
   - ☐ applied research
2. Would you also classify the nature of the research conducted by them as being *primarily* either
   - ☐ product research, or
   - ☐ process research
3. Please check any of the following categories that describe the literature facilities available to the “observed chemists.” You may check as many or as few of the categories as apply.
   - ☐ Current journals available to the men at their desks. (Personal subscriptions or routed copies.)

☐ Library facilities in the same building with the literature available for withdrawal.

☐ Main company library not in the same building, but literature can be withdrawn on request.

☐ Public library facilities available with a procedure for company use.

4. What is the average annual cost of the physical plant, *exclusive of the laboratory equipment and materials,* used by the "observed chemists?" Please include the amortization of the capital goods, any imputed rent, and operating and maintenance costs. $______/yr.

   What is the average annual cost of the laboratory equipment and supplies used by the "observed chemists?" Include any amortization of capital equipment in this figure. $______/yr.

   [If the "observed chemists" do not have separate facilities, apportion the costs on a manpower basis.]

5. This question refers to a group of people which includes the chemists who were observed for two weeks earlier this year. We want to include in this group most of the people who work and communicate with the "observed chemists." This will usually be all of the men in the laboratory, or all of the men in same building, but it can also include men from other units of the company, if they are in frequent contact with the "observed chemists."

Classification by specialty	*Number in group*	*Average annual salary*
Mathematicians and statisticians	______	
Physical scientists, including physicists and metallurgists	______	
Chemists, including both "observed chemists" and all others	______	$______ (for all groups above through Behavioral scientists)
Chemical engineers	______	
Other engineers	______	
Biologists, medical men, life scientists	______	
Behavioral scientists, including psychologists and economists	______	
Technicians, laboratory assistants	______	$______
Secretaries, clerical staff	______	$______
Administration personnel	______	$______
Total in group	______	

6. There were [ ] "observed chemists" in your group, and your observer has a list of their names. Please indicate here their average annual salary: $_______.

7. Please write here any comments you wish to make about this questionnaire:

____

____

____

____

____

Information and Literature Use in a Research and Development Organization

I.H.HOGG and J.ROLAND SMITH

ABSTRACT. A uniform sample drawn from three arbitrary status grades of applied scientists (Research Managers, Senior staff, and Junior staff), totalling 157 persons, were interviewed using standard questionnaires, and also were given 14-day reading diaries to complete. Chief information sought was: *(a)* how they obtained their scientific and technological information and how they valued the different sources; *(b)* how many abstracts, periodicals, research reports and textbooks they read during 14 consecutive days, and where they read them; *(c)* whether they considered they had adequate time for reading at work; *(d)* where they obtained the literature read during the 14 days, how they got references to it, and how much literature they bought themselves; *(e)* their criticisms of the various library lists as reference sources; *(f)* how they used the information gained during their 14 days' reading, and what reference-sources led to the most useful reading; *(g)* the value they placed upon periodicals according to age and language, and their use of those of British and foreign origin; *(h)* whether they kept personal data records; *(i)* their suggestions for new, and criticisms of existing library services; together with the formal qualifications and field of research of those interviewed. All the sample were interviewed, and 92% of the diaries were returned.

The first analysis of results showed that: *(a)* The prime information sources were informal contacts and the literature, of which reports were valued most and periodicals least by those answering. *(b)* Less than one abstract-consultation was made per head during the 14 days' reading; less than one-third of the sample read any abstracts during this period. Two-thirds of the consultations were for keeping up with the literature, one-third for locating past literature. All except one of the diarists read some periodicals, reports and textbooks, an average of 4 per head. Three-quarters of the reading was in working-hours, the rest at home. These figures are, however, statistically suspect, *(c)* Three-quarters of those interviewed said that for some part of the past working year they had no time to read in working

I.H.HOGG and J.ROLAND SMITH Industrial Group, United Kingdom Atomic Energy Authority, Risley, Warrington, Lancashire, England.

hours [cf. *(b)*]. *(d)* Over half the diarists' literature was from the Group libraries, one-quarter, mainly reports, was sent direct by authors or colleagues, one-seventh was the diarists' property, and a small amount was borrowed from colleagues. Less than half the sample bought their own books, one-third bought their own journals: average spending per head in the past year was £4 on books, £2 on journals. Of the 14 days' reading, no references were required for 40% of it; of the remainder, colleagues recommended 18%, the diarists' memory or knowledge accounted for 18%, the library provided references for 14%, references in other publications were 6%, and abstract journals (and the library catalogue for books) provided 4%. *(e)* Of those interviewed 25% criticised the library bulletin (of selected journal references), 13% book accessions list, and 17% the report list. All lists were used by about three-quarters of the sample. The main tendency was to ask for more selective lists for individuals or their immediate departments, and for more abstracts or annotations in the lists, *(f)* Half the diarists' reading was in aid of their research work, one-third for general interest, and only about 3% was discarded as of no use. The highest percentage of "discarded" per reference-source occurred when the least-used source, abstract journals, was used (about 14%). *(g)* With a maximum usefulness-score of 6, among those who used them current periodicals of research scored 4.6, falling after 10 years to 2.9; current periodicals of technology scored about the same. Decline of interest in both types of periodical was heaviest among engineers, least among chemists and metallurgists. Of foreign-language periodicals, among those who used them the "face-value" (including the effect of the language-bar) was lowest for the Japanese periodicals; the potential value (assuming no language-bar) was highest for the German; and the difference between face-value and potential value was highest for the Russian. Of periodicals of British and foreign *origin,* 57% of the scientists' information came from the former, and 42% from the latter, *(h)* Personal records of data or useful references were kept by two-thirds of the sample, and another one-tenth of them used records kept by others in their section, *(i)* Two-thirds of the sample offered criticisms of or suggestions for improvement of the library service: only those of wider interest are mentioned. Major comments were: *(a)* the libraries should publicise their services, *(b)* more or better qualified library staff are needed, *(c)* librarians should notify users individually of literature of interest, *(d)* better copying (reprint) facilities should be provided.

1. INTRODUCTORY

This paper records the major part of a survey of which the prime object was to provide information for use within the United Kingdom Atomic Energy Authority. In consequence, not all the data obtained is presented, but that judged to be of fairly wide interest has been included. It is not proposed here to review the surveys of the use of literature and information by scientists which have been conducted by others in the field: this has already been excellently done up to 1956 by R.R.Shaw (1). Nevertheless, in planning the present investigation and in writing this report due regard has been paid to the surveys of other workers, chief among them being those of J.D.Bernal (2), S.Herner (3), and R.G.Thorne (4), from which valuable help was obtained both in assessing fruitful areas for enquiry and in the actual design of the questionnaires.

The subject of the survey was the applied scientists and technologists in the Research and Development Branch of the U.K.A.E.A.'s Industrial Group establishments, who work at five laboratories and at the Group's headquarters. The Research and Development Branch is comparable with similar organisations in large-scale industry. It assists the Engineering (design) and Operations (management) Branches by *(a)* developing existing processes, *(b)* removing production difficulties, and *(c)* providing information for the design of new plant, or prototype nuclear reactors. Four of its laboratories are sited with large industrial plants scattered over the northern part of Great Britain; the fifth is near the Group headquarters at Risley. These laboratories have their own programmes of research and development, coordinated from Risley, where there is an advisory service. The works which they adjoin have, apart from wide geographic scattering, a common characteristic in their very recent origin (the first establishment dates only from 1946, and elsewhere construction still continues), and the highly sophisticated nature of the plant, in which technological progress is very rapid. There are, therefore, acute information communication problems to be met, on the score of *(a)* evolving technology, *(b)* lack of tradition, and *(c)* geography and partial autonomy. The method mainly employed is through technical liaison, design, and management committees on which R. & D. representatives sit, and to which reports are submitted for discussion. Report-writing is consequently a major R. & D. activity, and presents a major documentation problem, exacerbated to some extent by the access restrictions imposed by military security.

In addition to the internal communication problem, there is the more familiar one of efficiently communicating to R. & D. scientists the results of research by outside workers. This was recognised and dealt with from the beginning by establishing libraries, controlled by the R. & D. Branch, at each laboratory and at Risley (the libraries also serve the works staff), with coordination and certain common services to all the libraries by a chief officer at Risley, where the most highly developed library is situated. In recent years, however, the libraries have played an increasingly large auxiliary part in internal communication by receiving, publicising (subject to security), retrieving and lending internal reports. By virtue of their control by their principal users—the R. & D. Branch—and their decentralization, the libraries ought to be in a most favourable position to fulfil their role of getting the right information at the right time to the right person. Whether in fact they were successful in this was one of the main purposes of the present survey to find out. At the same time, it was plain that the investigation should try to discover what were the scientists' real needs in the sphere of communication and documentation, to gain a sound basis for library and other reforms. It was thought that this part of our survey could well be of interest to workers outside the Industrial Group, and that belief gave encouragement to a wider investigation than had originally been planned.

The main heads of information sought in the survey were as follows:

(a) How the scientists obtained their scientific and technological information and how they valued the different sources.
(b) How many abstracts, periodicals, research reports, and textbooks they read during 14 consecutive days, and where they read them.
(c) Whether they considered they had adequate time for reading at work.
(d) Where they obtained the literature read during the 14 days, how they got references to it, and how much literature they bought themselves.
(e) Their criticisms of the various library lists as reference sources.
(f) How they used the information gained during their 14 days' reading, and what reference-sources led to the most useful reading.
(g) The value they placed upon periodicals according to age and language, and their use of those of British and foreign origin.
(h) Whether they kept personal data records.
(i) Their suggestions for new, and criticisms of existing library services; together with their formal qualifications and field of research.

The method of survey was a combination of personal interviews based upon a standard questionnaire, and pro-forma “diaries.” This was decided upon because, first, there was little time and no special staff available for the application of refined techniques (such as R.R.Shaw's suggested supervised diaries *(op. cit.)* or for the application of operational research to the problem), and secondly, because much of the required data was of a detailed quantitative nature. It seemed plain that while interviews were a convenient and, for the interviewed person, a relatively agreeable and quick means of getting to know about their general practices and preferences, if interview questions were to be asked about, for example, from what sources they got the literature-references to what they had read during a specified period, their answers would probably be no more than inspired guesses. For this purpose a diary, or similar *record,* appeared more likely to yield realistic results, though it was known that it would not be 100% reliable. As, however, the survey was conducted under official auspices it was expected that cooperation would tend to be greater than under truly voluntary conditions.

Before beginning the main survey, a pilot survey was made of 21 scientists at two of the laboratories during September and October 1957. Only staff actually engaged upon or responsible for supervising research were considered; administrative or quasi-administrative personnel were excluded. One individual was selected at random from each of the U.K.A.E.A. staff status grades, in each of the two laboratory staff nominal rolls. A personal letter from one of the present authors was sent to each of those selected, explaining the purpose of the pilot and the main survey which would follow, asking for criticisms of both interview and diary, and guaranteeing that all records would be treated as entirely confidential (this promise has been kept). Response to the pilot was 100% and some useful suggestions were received: questionnaires and diaries were therefore amended, mainly by way of simplification, or by transfer of questions from interview to diary, and the final versions printed. That part of the final questionnaire which has been analysed for this paper, and the whole of the diary, are reproduced in Appendices I and II.

2. MAIN SURVEY: COMPOSITION OF THE SAMPLE

For the main survey, which took place between November 1957 and March 1958, several hundred individuals were available for investigation, distributed over the six Industrial Group establishments. From these, as in the pilot, those doing administrative or quasi-administrative work were deducted, together with those who took part in the pilot. The scientists on the six establishments' nominal rolls were then separately sorted into three arbitrary status grades: Grade 1 represented the Research Management, Grade 2 the Senior scientific, technological, and experimental personnel, and Grade 3 the Junior scientific, technological, and experimental staff. Names within each grade were then alphabetised and a random sample was taken within each grade (the proportion being the same in each grade), totalling 157 persons, for participation in the survey. Almost all those so selected took part, but where any individual was not available (e.g., because of illness or absence), the next name on the nominal roll was selected. The analysis of the results was done manually, the number involved hardly warranting the use of punched cards.

The composition of the whole sample according to status grade was as follows:

Grade 1	*Grade 2*	*Grade 3*	*Total*
Research Management	Senior staff	Junior staff	
6 (3.8% of whole sample)	57 (36.3%)	94 (59.9%)	157 (100%)

At each interview the scientists were asked to state the general field of applied research in which they were working (*vide* Question 4, Appendix I). Their answers were divided into five main classes, and the number of persons in each class was: biology, 2 (1.3% of sample); metallurgy (and metal physics), 32 (20.4%); engineering, including electrical engineering, 35 (22.3%); physics and mathematics, 39 (24.8%); chemistry and chemical engineering, 49 (31.2%). The total was 157 (100%).

The samples were then asked the type of formal qualifications they possessed (*vide* Question 2, Appendix I). Their answers were divided into *(a)* those with university degrees, *(b)* those without degrees but with a professional qualification *of a senior institution,* and *(c)* those without either of the foregoing. The result showed 62.4% with degrees, 15.3% with professional qualifications, and 22.3% with neither (but not necessarily no qualifications). It was intended to analyse part of the answers to questions according to the formal qualifications of the scientists, to discover whether a university education influenced their information habits and wants. A specimen analysis was made of Questions 7/14 of the questionnaire, but the results in that case were almost exactly typical of the whole sample, and analysis by formal qualifications was therefore abandoned.

3. ANALYSIS OF INTERVIEW QUESTIONS

3.01. QUESTIONNAIRES

Interviews (of about 45 minutes' duration) were conducted by a single interviewer, chosen from the junior scientific staff of Risley Library by the Group Training Department. The same interviewer was used for both the pilot and main surveys, and he was thoroughly briefed on both occasions by one of the present authors. The questionnaires used were provided with pre-coded answers to the majority of questions, partly to speed up marking by the interviewer, and also (in the case of Questions 7/14, 33/36, and 37 *vide* Appendix I) to standardise the scoring when the scientists were asked to evaluate various sources of information according to usefulness. When asking these evaluative questions, the interviewer handed the scientist a list from which to choose his answer. All those taking part were asked if this method gave them adequate expression, and so far as concerns that part of the survey reported here, it was considered satisfactory. These answers which ranged from "No use" to "Essential," were given numerical scores from 1 to 6. The average score was then calculated for each question, resulting in an average value for each information source. Sources could then be arranged in an order of usefulness. Herner *(op. cit.)* had a similar objective, but originally asked his subjects simply to arrange the sources in order (which would appear to have the effect of discouraging, if not actually preventing scientists indicating sources of equal usefulness). Because of this lack of independent evaluation, he evolved a corrective weighting based upon the amount of use they made of the literature. This proved to have substantially the same result as asking the scientists simply to say whether they used a source or not, counting "yes" as 1 and "no" as 0, and then totalling the scores for each source. He therefore abandoned the first in favour of the second method. The evaluation scoring used in the present survey appears to have secured fine gradation of usefulness *among those using* a source of information, i.e., when the scoring was not weighted to account for those not using the source. The validity of the results does however depend upon consistency of each individual in his use of the evaluation scale. Nevertheless, it must be admitted that when the scores were weighted to account for those not answering (*i.e.,* answers "Does not apply" and "Don't know" were included in the score) and were compared with the results of simple additive scoring as used by Herner, there was little difference between them. Both unweighted evaluation scoring and additive scoring have therefore been adopted, the former having its own special value.

3.02. QUESTIONS 5, 6: ADEQUACY OF TIME FOR READING AT WORK

Because the largest part of the survey concerned the use of literature (and in the diaries actually took note of what reading was done during a fourteen-day period), those taking part were asked whether they considered they had adequate time for reading during working hours. 73% said there was insufficient time; the remaining 27% found the time adequate. Significantly, 100% of the Research Managers said they had insufficient time, as did 81% of the Senior grade; even 67% of the Juniors felt the same. The 73% with inadequate reading time during working hours were then asked for how many weeks in the past year (excluding vacation leave) they were unable to read at work. 17% of them did not know, but the remaining 83% gave the following average figures, which, if any great value is placed upon the reading of scientific and technical literature, reveal a very unhealthy situation, most serious in the higher status grades:

Research Managers	35 weeks (76% of working year of 46 weeks)
Senior grade	29 weeks (63% of working year of 46 weeks)
Junior grade	25.5 weeks (55% of working year of 46 weeks)

The average was 27 weeks (58% of the working year).

3.03. QUESTIONS 7–14: EVALUATION OF SOURCES OF INFORMATION

It could be argued that lack of time for reading in working hours need not necessarily be a hindrance, if the scientists obtained their information by other means. They were therefore asked independently to evaluate a series of information sources, according to the usefulness of each for their research work. The average values placed upon each source by those who used it were obtained as described in *3.01* above. The results are given in Table 1, with the sources placed in order of usefulness *to those who used them*. For comparison, the numbers of those who offered a valuation of each of the sources of information are also given as percentages of the sample, from which their usefulness to the whole sample may be deduced.

TABLE 1

Source of information	*Average score of those answering*	*Per cent answering the question*
Consulting relevant reports, etc.	(5.1) Very useful	96
Contacts with others in your field	(5.0) “	98
Consulting relevant books	(4.9) “	99
Consulting relevant journals	(4.6) “	96
Attending official meetings of committees, etc.	(4.0) Useful	52
Consulting relevant committee minutes	(3.9) “	68
Actual information (*not* literature references) from the Library	(3.8) “	70
Attending external conferences and professional meetings	(3.8) “	47

All the questionnaires were then analysed according to the formal qualifications of the sample, but these were found to have no effect upon the answers given. Analysis according to the status grade showed that this had little influence upon the answers, except that 100% of the Research Managers rated “Contacts with others,” “Books,” “Reports,” and “Committee Minutes” higher than the other two Grades, and the two-thirds of them who answered rated “Information from the Library” lower than did the other two Grades.

To sum up, therefore, it was conclusively demonstrated that the literature occupied a virtually preeminent place as source of applied research information, regardless of either the status or formal training of those in the sample, a place shared only by the personal contacts with other workers in their field.

3.04. QUESTIONS 15/20: LITERATURE SEARCHING HABITS

The scientists were next asked how they usually obtained their references to the literature they read—whether for a particular paper, or for a collection of references on some particular subject. The main object was to discover how far the library services were releasing the scientists from this work, and if they were not, what requirements were not being met by the libraries. Asked whether they always personally carried out a search without the help of others, 24% said they always searched personally, 8% never searched personally, and 66% said they sometimes did their own searches. Only 1% did not do searches or find them necessary. These replies were analysed according to the status grade of the scientists (Table 2).

TABLE 2

Per cent of status grade who:	*Research Management grade, %*	*Senior grade, %*	*Junior grade, %*
Always do their own literature searches	0	14	31
Sometimes do their own literature searches	33	75	64
Never do their own literature searches	67	9	3
Never do or need literature searches	0	2	2

Those who *sometimes* did their own searches (66%) were asked if they delegated the work to their staff; 70% of them said either that they had no staff or that they never delegated searches (and must therefore ask the Librarian occasionally); 30% of them did delegate to their staff, but (except for a small number) they also approached the Librarian themselves on occasion. Asked if, when their staff were delegated the searches, they sought library help, 28% replied “always,” 53% said “sometimes,” 13% “never,” and 6% “don't know.”

The scientists who *always* conducted their own searches (24%) were then asked why they did not call upon library assistance; 47% of their answers were definitely critical of the library or contained implied criticisms of it; 2% did not know of this library service, and the remaining 50% either preferred to do searches themselves or said they had their own indexes.

Of those who *never* did their own literature searches (8%), 50% said they always asked the Librarian, 42% sometimes asked the Librarian, sometimes delegated searches to their staff.

Finally, those who made use of their Librarian at all for literature searching (about 70%) were asked why they did so. 42% of them said the Librarian knew the sources better, or was best qualified to search; 40% of them said the library saved their time; and 14% said "because the Librarian had the sources at hand."

Summarizing and referring to Fig. 1, it would appear, therefore, that, apart from *(a)* 24% of the scientists who for one reason or another always found their own literature references, *(b)* 4% who sometimes found their own but never asked the Librarian to help, *(c)* 1% who did not do literature searches or find them necessary, the Librarian is consulted by all scientists on some occasions, and generally also by their staffs when the work of literature searching is delegated to them. The people who always did their own literature searching (24%) included no Research Managers, and consisted almost wholly of Junior grade staff: only 21% of them (8 persons) were at Senior grade level.

3.05. QUESTIONS 21, 22: PERSONAL INDEXES

Enquiry having been made into their literature searching habits, it seemed worth discovering how systematic the sample were in using data obtained. They were therefore asked whether they attempted to keep their own indexes or notes of data supported by references, or of useful references; or whether they depended on someone else's index. Sixty-six per cent of the sample said they maintained their own, 5% intended to do so, 10% did not keep their own but depended on someone else in their section, and 19% neither kept their own nor depended on one in their section. An analysis of those who kept their own indexes is given here. According to their status grade the percentages were: of Research Management grade, 33%; of Senior grade, 83%; of Junior grade, 59%. The percentage for the whole sample was 66%. According to the general field of their applied research work the percentages were: of biology, (2 persons), 100%; of metallurgy and metal physics, 69%; of engineering (including electrical engineering), 46%; of physics and mathematics, 72%; of chemistry and chemical engineering, 74%. The percentage for the whole sample was 66%.

The low figure for the engineers suggests that either such data as they need can readily be obtained from a few handbooks, or that they carry more of it in their heads than do their colleagues.

3.06. QUESTIONS 23/28: CRITICISMS OF LIBRARY LISTS

Because the Headquarters Library of the Industrial Group compiles and regularly issues three kinds of literature lists (accessions or selections) as aids to the choice of literature references by staff throughout the Group, the scientists who either received personal copies or used office copies of these library lists were invited to comment critically upon them. The lists concerned are fairly typical of the "comprehensive" type used in large organisations, and the present

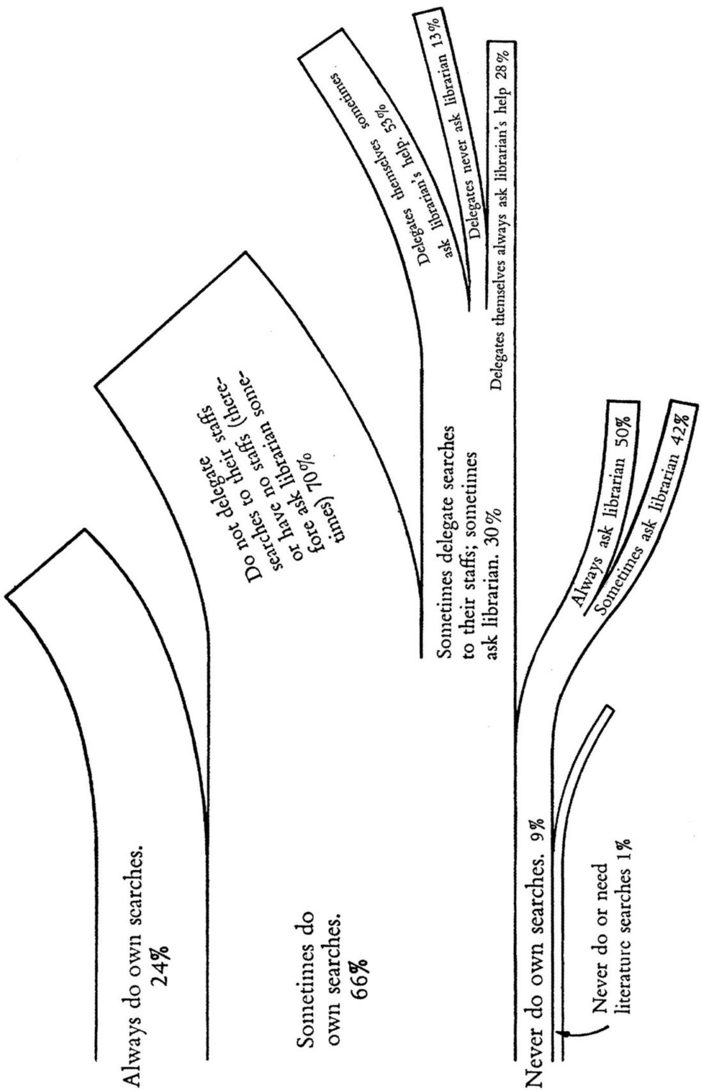


FIGURE 1. Questions 15/20—literature searching habits.

writers felt that at least some of the users' criticisms would have interest outside the Group. They consist of *(a)* a weekly bulletin of articles selected by the H.Q. library (with contributions from the outstation libraries) from journals, etc., arranged under broad subject headings, for the most part without abstracts (to delay publication as little as possible), together with news of meetings, etc., and a list of press releases; *(b)* a fortnightly accessions list of books received, also under broad subject headings, without abstracts; *(c)* a weekly accessions list of British reports, listed in groups under their originators, with an appendix of U.S.A. reports listed under broad subject headings. Only the Reports List contains "classified" information, but it is fairly readily available to Industrial Group scientific and technological staff.

It was found that the usership of the three publications ranged from 69% to 76%. Criticisms of the "Bulletin" were offered by 25% of its users, of the Books Accessions List by 13%, and of the Reports List by 17%. The major critics were the metallurgists, followed in order by the biologists, the chemists, the engineers, and lastly the physicists.

Although the small number of critics might be said to indicate that on the whole there is nothing much wrong with these publications, some of the comments made below deserve consideration.

All publications:

Abstracts (or short reviews) should be provided. [Occasional abstracts, or annotations, are already given.]

A more specialized, less general, arrangement of the contents is wanted. [Some critics wanted a selection specially for their departments.]

Does not adequately cover his subject-field. [Mainly the chemists said this.]

The Bulletin:

Insufficient coverage of foreign periodicals. [One metallurgist and one chemist.]

His copy should be marked specially to show articles of interest to him. [Two engineers.]

Straightforward contents-lists under each journal title are preferred to the present individual-article arrangement under subject-headings. [Two metallurgists.]

The Books Accessions List:

Should also list new books of interest, which have not been added to the Library. [One engineer and one metallurgist.]

A cumulative index should be published. [One engineer. In fact the libraries' union card catalogues provide this.]

The Reports List:

The reports should be listed under subject headings rather than collected under names of originating bodies. [Two metallurgists and one physicist. In fact, when the number of reports warrants it, i.e., in the case of U.S.A. reports, subject headings are already provided.]

From the above it is plain that there are some conflicting requirements, the small number of which suggests that they should be ignored. The desire for more abstracts, and the requests for more specialized lists, or lists specially marked for individuals, all point to the need still further to sharpen selection and thereby to save the time of users.

3.07. QUESTIONS 39, 40: CRITICISMS OF THE LIBRARY SERVICE

Further criticisms of and suggestions on the library service as a whole were asked for from the sample. Sixty-five percent (102 persons) offered some comments, many of which were of purely domestic interest, but the more significant of which are given below:

The Library should issue a list of their various services. [17 persons.]

His Library has insufficient, or insufficient qualified Library staff. [16 persons.]

The Library should notify him of current reports and periodical articles in his field, as they are received. [14 persons.]

A better and quicker photocopying service [for production of reprints] is necessary. [10 persons.]

The translation service should be improved. [Russian, 2 persons; other languages, 4 persons.]

The Library should compile technical and design-data manuals on U.K.A.E.A. subjects. [4 persons.]

A list of Library bibliographies should be circulated periodically. [4 persons.]

More, and more up-to-date bibliographies should be compiled by the Library. [4 persons.]

Literature surveys ought to be produced by the Library. [3 persons.]

The Library should record the work done by individuals, for reference by others. [1 person.]

Instruction ought to be given on the most efficient method of putting enquiries to library staff. [1 person.]

3.08. QUESTIONS 29/32: PRIVATE PROVISION OF LITERATURE

The sample were next asked if in the past year they (1) bought any of their own books, and if so, how much they spent, (2) subscribed (other than through professional membership) to any journals, and if so, how much these cost. It was found that 48% bought their own books, and 32% subscribed to their own journals. Average expenditure per head by these people was £4.3 on books and £2.4 on journals (after neglecting the answer of one Research Manager who spent £125 on books.) A further analysis by status grades is tabulated in Table 3.

TABLE 3

	Books bought		*Journals subscribed*	
Status grade	*Bought own, %*	*Expenditure per head*	*Subscribed, %*	*Expenditure per head*
Research Managers	50	£14[a]	50	£6.3
Seniors	45.6	£3.7	40.3	£2.6
Juniors	50	£4.2	25.5	£1.8

[a] Neglecting one Research Manager who spent £125.

Though it is evident, and not surprising, that the Research Managers spent per head considerably more than the other two grades together, it is interesting that the Senior and Junior grades spent about the same, with much the same proportions, i.e., about 50%, doing the spending.

3.09. QUESTIONS 33/36: VALUE OF PERIODICALS ACCORDING TO AGE

The value of current and older periodicals as a source of information was next enquired. The sample were asked to assess the usefulness, according to their age, of journals in their spheres of interest devoted *(a)* to research and *(b)* to technology and industrial news. The evaluation of usefulness was made and scored by the same means as that used for Questions 7/14 *(vide* 3.03) resulting in independent scores for each case. The percentages of those answering are shown together with these, in Table 4.

TABLE 4 Usefulness of periodicals by age

	Journals of research		*Journals of technology and industrial news*	
Age, years	*Average evaluation score of those answering*	*Per cent of sample answering*	*Average evaluation score of those answering*	*Per cent of sample answering*
Less than 1	(4.6) Very useful	82	(4.7) Very useful	84
From 1–5	(4.4) Useful	81	(4.2) Useful	80
From 5–10	(3.5) Fairly useful/useful	81	(3.1) Fairly useful	81
Over 10	(2.9) Fairly useful	79	(2.5) Little use/ fairly useful	79

The questionnaires were further analysed according to the field of research of those answering the questions; this analysis is set out in Fig. 2.

If the biologists are excepted, as unlikely to have much concern with technological journals, it is apparent that the aggregate decline of interest in journals

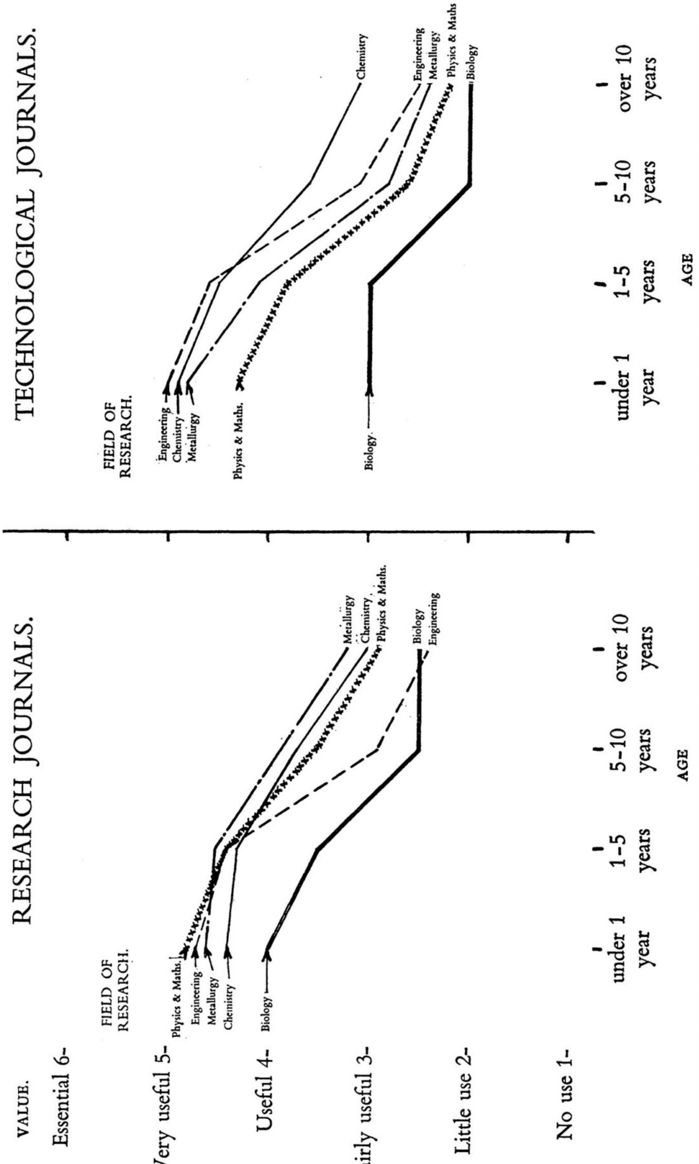


FIGURE 2. Questions 33/36—value of periodicals by age; average evaluation by those answering, according to their field of research.

of technology was least among the chemists (1.8 difference) followed, by the physicists and mathematicians (2.1 difference), the metallurgists (with 2.4 difference), and the engineers (2.5 difference). Of the journals devoted to research, the least aggregate decline of interest was shown to be shared between the chemists and the metallurgists (each with 1.4 difference), followed by the biologists (1.5), physicists and mathematicians (1.9) and again, lastly, by the engineers (2.3 difference).

3.10. QUESTION 38: DOMESTIC AND FOREIGN PERIODICALS, PROPORTIONATE USE

The foregoing enquiry was followed up with one about the percentage of research information obtained from British as against foreign periodicals, again dividing them into journals of research and those of technology and industrial news. The question was answered by 83% of the Research Managers, 86% of the Seniors, and 76% of the Juniors, or 80% of the whole sample. Again the replies were analysed according to the field of research of those concerned (Table 5).

TABLE 5 Percentage of information obtained from British and Foreign periodicals

	Journals of research		*Journals of technology*	
Field of research of those answering	*British, %*	*Foreign, %*	*British, %*	*Foreign, %*
Biology	40	60	Don't know	Don't know
Chemistry and chemical engineering	54	46	46	54
Engineering	61	39	67	33
Metallurgy	57	43	57	43
Physics and mathematics	61	39	60	40
All fields	57.5	42.5	57.2	42.8

It is apparent that according to the average opinion there is only a slightly greater usage of British as against foreign journals, with a ratio of 1.3:1, and no significant difference in this ratio as between research and technological periodicals. The greatest use of British research periodicals was made equally by the engineers, and the physicists and mathematicians; the greatest use of the foreign ones by the biologists. Of technological journals, the engineers made most use of the British, and the chemists made most use of the foreign ones.

3.11. QUESTION 37: VALUE OF PERIODICALS ACCORDING TO LANGUAGE

Probing more deeply into their use of periodicals, the sample were asked to assess the usefulness of those in their field of research, independently according to the language in which they were published. It was realised that the language-barrier would probably introduce an undesirable bias into an individual's valuation of the literature of certain languages, and the question was consequently designed to take account of this and to illustrate the effect of the language-barrier. The sample were therefore asked to state: *(a)* the face-value (including the effect of language-barrier), *(b)* the potential value (assuming no language difficulty, or easily available translations). All answers were volunteered, that is, a list of languages was not recited by the interviewer, in the hope that the scientists' views would be from their experience. Fifty-eight per cent of the sample offered a valuation in one or more foreign languages, having either opinions or experience, the 42% remaining having either no opinions or making no use of foreign-language periodicals. These 42% were made up of one-quarter of the whole Senior status grade and over half of the Junior grade, a somewhat disquieting state of affairs.

Scoring of the scientists' evaluations was as in Questions 7/14 (*vide* 3.03), i.e., as averages of those answering (because so few people evaluated some languages, any attempt to account for those not answering would have concealed the usefulness to those who did value them). In the case of individuals who said they experienced no difficulty with particular languages, this fact was recorded in lieu of the potential valuation they should otherwise have given, and the latter was assumed to be the same as the face-value. The only cases of this sort occurred with the French, German, and Russian language periodicals where, respectively, 14%, 5%, and 1% of all the answers were that there was no language difficulty. This, again, gives cause for considerable concern over the apparent narrowness of the scientists' reading.

The analysis of results is shown in Fig. 3. It will be noted that the evaluation scores here show considerable differences compared with the percentage of the sample answering. The differences between face-value and potential value are shown to be lowest for the French literature (one point), and highest for the Russian (2.8 points). German literature is rated as the highest in potential value, followed jointly by Russian and Japanese.

In an analysis of potential value according to the field of research of those answering the question, there were variations of only one point up or down from the average score with the single exception of one engineer who rated Japanese literature as potentially (2) “Little use.” The remainder of these results were therefore discarded.

4. ANALYSIS OF READING DIARIES

4.01. CONDITIONS OF THE TEST

The diary provided for a consecutive fourteen-day record (i.e., both in and outside working hours) of the amount and kind of scientific or technological literature read, to be indicated at the time of reading (see Appendix II for complete

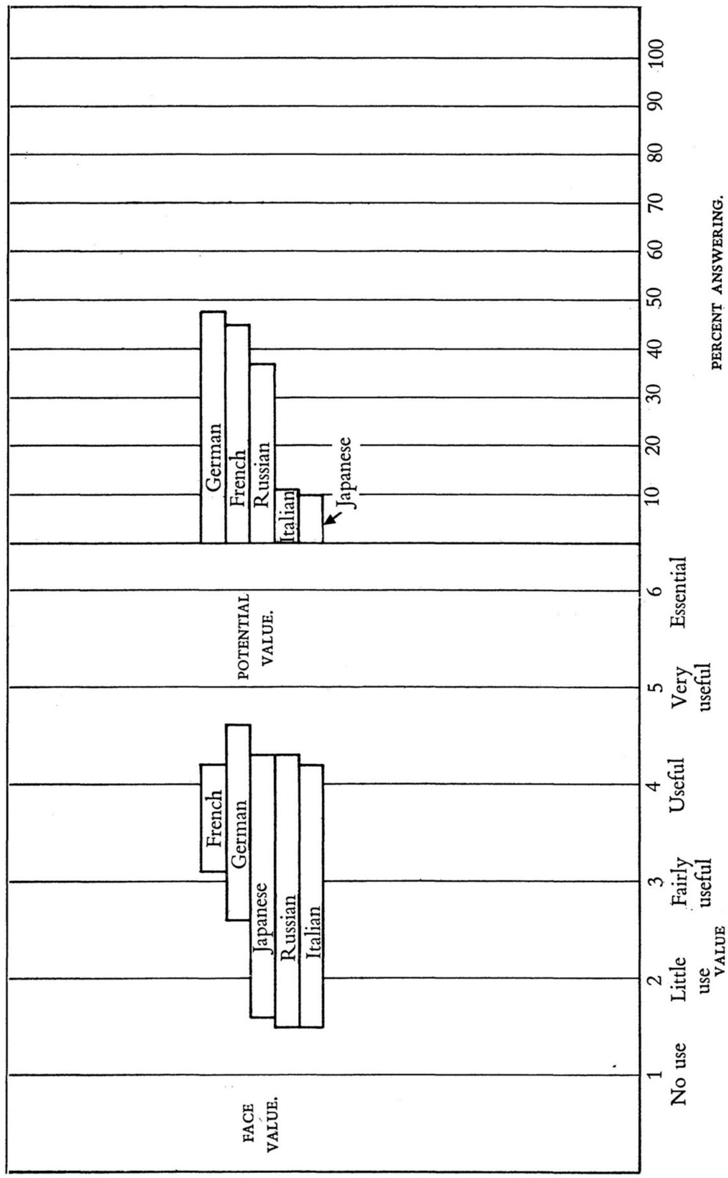


FIGURE 3. Question 37—value of periodicals according to language.

diary). Four types of literature were provided for: abstracts, periodicals, research reports, and textbooks, with a separate diary sheet for each, made up into pocketable booklets 5 by 7 inches. As in the diaries used by Bernal (*op. cit.*), distinction was made between "scanning" and "reading" each kind of literature (abstracts excepted), but the reader was asked to indicate which applied by checking the appropriate column on the diary sheet. Unlike Bernal's diaries, provision for identifying the literature by name was only made for abstracts (by checking a printed list of titles). Here two check marks could be used: one to show that the abstract journal was being used to locate or identify-past literature, the other when it was being used to keep abreast of recent developments. The use of periodicals, reports and books was shown by checking horizontally in a series of vertical columns: the first divided into "scanned" or "read," and the remainder labelled "Where read," "Where obtained," "How reference to it was found," and "Use made of it," within each of which there were pre-printed answers. An additional column (on the periodicals sheet only) showed the number of actual issues of periodicals as well as the number of articles in them which were scanned or read (reprints or photocopies of articles were checked in both columns.)

Return of the completed diaries for analysis was prolonged much beyond the estimated period, and in fact the last were received only in early March. They had been sent out in batches at different times to different establishments to obtain a "spread" of reading period, but perhaps because participants were not given a fixed date on which to begin their record, they tended to put it off. They were, however, specially warned to start recording immediately, subject only to sickness or vacation leave, and regardless of whether they were able to do any reading or not. Despite this, *only 1 individual recorded no reading at all,* and it was evident from remarks made during some of the interviews, that many had postponed diary records until they were free to do some reading. 144 diaries (92%) were returned in time for analysis and found to be usable, though in a small number of cases either directions had been misunderstood or some records were incomplete; these scientists were therefore asked to undertake a second test. It was noted that the incomplete records usually featured "How reference to it was found," and indeed it is possible that many of the other scientists' answers in this column are inaccurate because of the difficulty of remembering. The overall accuracy depended largely upon the diarists completing their records *at the time of reading* (which they were specially asked to do), but the neatness of many records gave rise to suspicions that they were marked up afterwards, or that scientists are tidier workers than is often supposed!

Because of the late receipt of the diaries for analysis and inclusion in the present paper, only the preliminary results are given.

4.02. AMOUNT OF LITERATURE READ, AND WHERE READ

Because of the apparent deliberate choice by many of those taking part of a period when some reading was possible (and the natural tendency for this to be during working hours), the amount of reading and the place where it was done cannot be taken as typical. All those who took part in the test read some periodicals, some reports and some textbooks. (Abstracts are dealt with in 4.03.) In the total of 579 periodicals, of 750 articles, 48.5% were scanned and 51.5% were read. This was 4.0 periodicals and 5.2 articles per head. Of 756 reports, 41% were scanned and 59% read to make 5.3 reports per head. Of 569 textbooks, 67% were scanned and 33% read—4.0 textbooks per head.

The usage of the three kinds of literature per head is seen to be about the same, but though the number of periodical articles scanned was virtually the same as the number read, reports were more carefully read, by comparison. Only consultation of textbooks showed a preponderance of scanning, suggesting that these were used mainly for checking facts or references.

The places where reading was done, with the proportion of reading in each place of each kind of literature are given in Table 6. From the table it is plain that the claim (*vide* 3.02) that for an average of 58% of the past working year no time for reading was available in working hours cannot be reconciled with the conditions established from the diary analysis, but as has already been pointed out, the latter are believed to be unreliable as to quantity of reading and place where reading was done.

TABLE 6

Category	*Library, %*	*Office or laboratory, %*	*Travelling on duty, %*	*At home, etc., %*
Periodicals	26	48	2	24
Reports	5	72	3	20
Textbooks	12	52	2	34
All literature	14	58	2	26

4.03. READING OF ABSTRACT JOURNALS

One of the most striking features of this survey was the evidence of the small amount of use of abstract journals, amounting to only 0.9 *consultations* per head. Only 32% of the scientists consulted any abstract journal at all. In the absence of further information it is not possible to reach firm conclusions on why this should be so, but it may be because few of the libraries *circulate* abstracts (keeping them principally for use in the library), or alternatively that the Group libraries' own (selected) weekly bulletin, and book and report accession lists are found easier to scan for current reading references, and much more up to date than any abstract journal known. It is possible to find confirmation of the latter in the check of source of literature references in 4.05 below. The additional fact that of the 129 abstract consultations made, 64% were for keeping up with the current literature (the remainder were for locating past literature) nevertheless points to their continued usefulness to some people for this purpose. An examination of the use of abstracts by people in the five fields of applied research showed the biologists (two persons only) as greatest users with 3.5 consultations per head; the chemists and the metallurgists both with 1 per head, followed by the engineers with 0.9, and the physicists and mathematicians with 0.7. Largest use of abstracts as a means of keeping up with current literature came from the physicists and mathematicians (85% of their use) and the engineers (83%); followed by the biologists (71%), the metallurgists (58%) and the chemists (with 45%).

Table 7 shows the titles of Abstract journals consulted, with the number of consultations according to the fields of research of the diarists.

4.04. WHERE THE LITERATURE WAS OBTAINED

Those taking part in the survey were asked where they obtained each item they read. The possible answers were pre-printed in the diary, but have been compressed for the analysis to show four sources: *the library* (by circulation, direct, or on permanent loan), *the initiative of a senior or colleague* (passed on, lent, or circulated), *requested from a senior or colleague* (on loan), or *his own property*. A fifth column was headed "From other source."

As can be seen from Fig. 4, the library was, as expected, the major source.

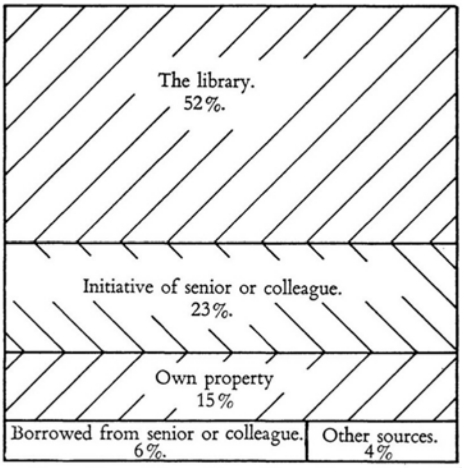


FIGURE 4. Reading diaries—where the literature was obtained.

The large figure for "initiative of seniors or colleagues" is mainly attributable to reports, of which a very large number are normally sent by authors direct to all those concerned in a project (quite independently of the library): these did, in fact, amount to 19% of all the literature read.

TABLE 7

	Biology	*Chemistry and chemical engineering*	*Engineering*	*Metallurgy*	*Physics and mathematics*	*Total*
Analytical abstracts		11				11
Annual review of metals literature				1		1
Bibliography of industrial diamond applications [Industrial diamond abstracts]				1		1
Biological Abstracts	1					1
British ceramic abstracts			1			1
British Hydromechanics Research Assn. bulletin			2			2
British Non-Ferrous Metals Research Assoc., Bulletin			1	3	1	5
Ceramic Abstracts		1				1
Chemical Abstracts	3	11		1	1	16
Corrosion		2		1		3
Electronic and Radio Engineer			2			2
Engineering Index			2			2
Engineer's Digest		1	1	2		4
Index to literature on spectrochemical analysis		1				1
Iron and Steel Institute, Journal		3	1	4	2	10
Journal of Applied Chemistry	1					1
Light Metals Bulletin			1			1
Metallurgical Abstracts				9	4	13
Nickel Bulletin			4	3	1	8
Nuclear Science Abstracts	2	8	5	1	6	22
Production Engineering Research Assoc. Bulletin			5	2	3	10
Physics Abstracts		3	1	2	2	8
Titanium Abstracts			2	1		3
Vacuum			2			2

4.05. HOW THE REFERENCE TO IT WAS FOUND

Following the source of the literature itself, the diarists were asked how they obtained the reference to it (this order was adopted because some literature was received by routes, involving no actual "reference"). Some of the pre-printed answers are again compressed for this analysis as follows: *recommendation* to read, by author, senior, or colleague; *memory,* chance, or inside knowledge of the diarist; *the library,* either through its various lists or on the librarian's initiative; reference found in *another book, journal,* etc.; found in *abstract journal* or (textbooks only) in the library card catalogue; or *no reference involved.* As will be seen by reference to Fig. 5, the greatest proportion of reading did in fact appear in the "No reference involved" column, to which 40% of the literature was attributed. As, however, 44% of the literature *itself* was sent either on seniors' or colleagues' initiative or came from the library *(periodicals by circulation),* or was the diarist's private property (23%, 6%, and 15% respectively, *vide* 4.04), it would appear that the "no reference involved" figure should have been even higher: allowance must, however, be made for some confusion of marking between the two columns "Where literature obtained" and "How reference to it was found."

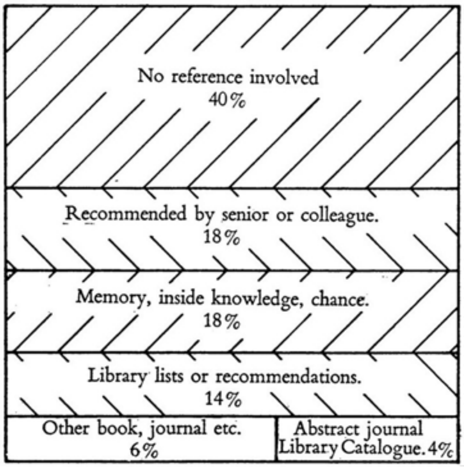


FIGURE 5. Reading diaries—how the literature reference was found.

The small proportion of references obtained from abstract journals (sharing lowest place with the library card catalogue) confirms the record of abstracts used, already described in 4.03 above. The library as source of reference appears remarkably insignificant, though its influence is likely to be a concealed factor in the case of "memory" of the diarist and "recommendation" by the diarist's colleagues.

4.06. USE MADE OF THE LITERATURE

The last record to be made by diarists concerned the use they were making, or intended to make, of what they were reading. The seven pre-printed answers have been compressed to five: *(a)* for your current or future research commitments, 48% of answers; *(b)* for general interest, 33% of answers; *(c)* for writing a report, article, book, lecture, 12% of answers; *(d)* information provided for a colleague or junior, 4% of answers; *(e)* discarded—not of use, or seen too late to be useful, 3% of answers.

The authors appreciated, when drawing up these pre-printed answers, that some of the diarists would not find it easy consistently to observe the distinction between answers *(a)* and *(b)* when actually recording what they read, however clearly they had established what that distinction was. It was considered that both the practical and the semantic difficulties of the two answers, particularly *(b),* were likely to produce a considerable degree of randomness in the aggregate results, and though as answers they seemed quite valid, it was nevertheless decided that it would be rash to draw firm conclusions, based on distinction between them. It was, however, noted that of the 33% reading "for general interest," two-thirds of it consisted of articles from periodicals, the reading of which, by the very nature of a periodical, would be likely to bring more incidental or extraneous matter under notice than is the case with other forms of publication. Of the 48% reading "for your current or future research commitments," (1003 literature items) the highest proportion was of reports (41%), followed by textbooks (36%) and periodicals (23%). It remains only to point to the very small amount of reading-matter discarded *(e)* as "not of use" (3%, or 70 items). Here again, views differ as to what is "useful," and it might be taking too sanguine a view to conclude, for instance, that the scientists' reading-habits were remarkably well-directed. Without, therefore, necessarily implying that answers *(e)* were interpreted by diarists as "absolutely irrelevant or useless reading," it was found that the highest percentage of *(e) per reference-source* occurred when the least-used source applied, i.e., abstract journals. These percentages were: textbooks, 17% discarded; reports, 14% discarded; periodicals, 11% discarded.

ACKNOWLEDGMENTS

The authors are indebted to Miss Kathleen Gales, of the Statistical Research Unit, University of London School of Economics, for valuable advice both in phrasing of the questionnaire and for suggesting the diary method for obtaining certain of the information; to Mr. R.O.Hughes, an operational research worker at the Springfields Laboratories of the Industrial Group, for statistical advice.on scoring of the questionnaire and phrasing of the questionnaire and the diary, and for undertaking a large part of the laborious work of analysis of completed results; to Mr. P.B.Woods of the authors' staff, who conducted all the interviews and collaborated in the analysis of results; to the Chief Scientists at the Risley headquarters for sanction to undertake the survey, together with the Heads of the five Research and Development Laboratories of the Industrial Group for permission for their staffs to be anonymously selected for collaboration in the survey (and particularly to Head of Laboratories, Springfields, for releasing Mr. Hughes and his staff to assist); to the 157 scientists and technologists who took part in the survey; and to the Managing Director, Industrial Group, and the Director of Research and Development for permission to publish it.

REFERENCES

1. SHAW, R.R. Pilot study on the use of scientific literature by scientists. U.S. Department of Agriculture, Washington, 1956 (unpublished).
2. BERNAL, J.D. Preliminary analysis of pilot questionnaire on the use of scientific literature. In Royal Society Scientific Information Conference, 1948. Report and papers submitted, paper no. 46, pp. 589–637.
3. HERNER, S. Information gathering habits of workers in pure and applied science. In *Industrial and Engineering Chemistry, 46,* 228–236 (1954).
4. THORNE, R.G. A survey of the reading habits of scientific and technical staff at the Royal Aircraft Establishment. Farnborough, R.A.E., 1954. 10pp. (Library Memorandum no. 16) (unpublished).

APPENDIX I SCORING SCALE FOR QUESTIONNAIRE

Apply to any question

X=does not apply (not used, etc.)

O=don't know

In Question 37

X="potential" value does not differ (because of no difficulty with the language)

O=no comment (foreign language periodicals not used)

Apply only to questions 7/14, 33/36, and 37

6=essential

5=very useful

4=useful

3=fairly useful

2=little use

1=no use

IN CONFIDENCE

Serial No. __________ ____________________ Date

R. & D.B. SURVEY OF INFORMATION COMMUNICATION
INTERVIEW QUESTIONNAIRE

I. *Personal Details of Interviewee*

1. Status grade [three arbitrary groups of actual grades]

(I) A Band / B Band / C Band

(II) P.S.O. C.E.O. C.I E.I M.I P.I / S.S.O. S.E.O. C.II E.II M.II P.II

(III) S.O. E.O. C.III M.III / A.E.O. E.III P.III

2. Degrees, professional qualifications
3. General field of formal training
4. Present general field of applied research (briefly)

II. *Information Communication in Research*

		Answer	
5.	Assuming that you do some scientific or technical reading during working hours, are you often without adequate time for it? [If NO, or X or O, ring (6) with "X" and go to (7)]	X YES	O NO
6.	For about how many weeks in the past year were you unable to do any reading during working hours?	X	O
7/14	Speaking generally, in order to obtain actual scientific or technical information to help you with your research work, how useful do you find the following:		
(7)	Attending official meetings (Committees, etc.)	X 6 5 4	O 3 2 1
(8)	Attending external conferences and professional meetings	X 6 5 4	O 3 2 1
(9)	Contacts with others (both in and outside UKAEA) in your field	X 6 5 4	O 3 2 1
(10)	Actual information (not literature references) from the Library or Information service	X 6 5 4	O 3 2 1
(11)	Consulting relevant books	X 6 5 4	O 3 2 1
(12)	Consulting relevant journals	X 6 5 4	O 3 2 1
(13)	Consulting relevant Committee Minutes	X 6 5 4	O 3 2 1
(14)	Consulting relevant reports and Committee technical papers	X 6 5 4	O 3 2 1

		Answer
15/20	Will you please now give some information on how you usually go about making a search for specific or suitable literature on a subject (this does not *necessarily* mean a "literature survey"):	
(15)	Firstly, do you do it personally (without calling on your staff or the Library or Information service to help)? [If NEVER or SOMETIMES, continue with (16). If ALWAYS, ring (16) to (19) with "X" and go to (20).]	X O ALWAYS SOMETIMES NEVER
(16)	Do you delegate the work to one of your staff (apart from the Librarian (or Information Officer))? [If ALWAYS, ring (18), (19) with "X," and continue with (17) and (20).] If SOMETIMES, continue with (17). If NEVER, ring (17) and (20) with "X" and continue with (18).]	X O ALWAYS SOMETIMES NEVER
(17)	When your staff do it for you, can you say if they generally get help from the Library (or Information service) in the searching?	X O ALWAYS SOMETIMES NEVER
(18)	Do you ask your Librarian (or Information Officer) to do the job for you? [If ALWAYS or SOMETIMES, ring (20) with "X," and continue with (19). If NEVER, ring (19) with "X" and continue with (20).]	X O ALWAYS SOMETIMES NEVER
(19)	Please say, briefly, why you ask your Librarian (or Information Officer) to do it.	X O
(20)	Will you say briefly why you never ask the Librarian (or Information Officer) to search for specific or suitable literature?	X O
21/22	We believe that a number of people maintain an index or record of scientific or technical data	
(21)	Do you keep a personal index or record of this kind, or a list of literature references containing valuable data? [If YES or INTEND TO ring (22) with "X."]	YES NO INTEND TO
(22)	Is this because an index or record is kept by someone else in your section?	X YES NO
23.	If you use the "Information Bulletin," have you any criticisms to offer? [If not used, ring 23, 24 with "X," and continue with 25. If NO, ring 24 with "X" and continue with 25.]	X O YES NO
24.	What are your criticisms, briefly?	X
25.	If you use the "Book List," have you any criticisms to offer? [If not used, ring 25, 26 with "X" and continue with 27. If NO, ring 26 with "X" and continue with 27.]	X YES NO

		Answer	
26.	What are your criticisms, briefly?	X	
27.	If you use "Reports List A" have you any criticisms to offer?	X YES	NO
	[If not used, ring 27, 28 with "X" and continue with 29. If NO, ring 28 with "X" and continue with 29.]		
28.	What are your criticisms, briefly?	X	
29.	During the past year, have you bought any of your own scientific and technical books?	YES	NO
	[If NO, ring 30 with "X" and continue with 31.]		
30.	About how much did you spend?	X	
31.	During the past year, have you subscribed to any scientific or technical journals (apart from those received under professional memberships)?	YES	NO
	[If NO, ring 32 with "X" and continue with 33.]		
32.	About how much did you spend?	X	

33/36	Turning to periodicals in general, will you say what is your experience of the usefulness of back-files of those in your field of interest, dividing them into: (i) journals primarily devoted to original research, and (ii) journals primarily devoted to technology and industrial news	(i)	(ii)
(33)	when less than a year old	O 6 5 4 3 2 1	O 6 5 4 3 2 1
(34)	from one to five years old	X O 6 5 4 3 2 1	X O 6 5 4 3 2 1
(35)	five to ten years old	X O 6 5 4 3 2 1	X O 6 5 4 3 2 1
(36)	over ten years old	X O 6 5 4 3 2 1	X O 6 5 4 3 2 1
37.	We would now like your assessment of the usefulness to you of foreign-language periodicals in your field of interest. However, because you may not know some languages very well, or even at all, you may judge the literature of those languages as of low value to you. In those cases we would therefore like your opinion on their POTENTIAL usefulness (assuming for example that translations were easily available). Will you therefore say how useful you find foreign language periodicals in your subject field, language by language, and each time give:	(a)	(b)
	(*a*) their actual usefulness to you,	O 6 5 4 3 2 1	X O 6 5 4 3 2 1

		Answer	
	(*b*) their potential usefulness (this does not, of course, apply if you read the language easily)		
38.	Would you say about what percentage of your research information is obtained from British as against foreign periodicals? Please deal separately with	(a) BRITISH % FOREIGN %	(b) BRITISH % FOREIGN %
	(*a*) periodicals mainly devoted to original research		
	(*b*) periodicals mainly devoted to technology and industrial news		
39.	Is there any SERVICE which your Library (and Information Service) is not giving, which you consider they ought to give you? (please specify briefly)	X YES	O NO
40.	Have you any final comments on or criticisms of your Library (and Information Service)? (please specify briefly)—	X YES	O NO

APPENDIX II READING DIARY

IN CONFIDENCE

R. and D. B. SURVEY OF INFORMATION COMMUNICATION

PART I: *Fourteen-day diary of scientific and technical reading.*

NOTES FOR COLLABORATORS

This project entirely depends on the conscientious and realistic recording of the reading you do throughout the 14 days' sample period. This means making the record at the time of reading, rather than from memory later on. It will certainly be inconvenient to do this, but you can reduce the inconvenience by careful study of the questions and the methods of recording answers, *before* you actually start. Apart from offering this advice, we can only invoke your patient co-operation.

Will you **please note the following points**, which are vital to the Survey :—

1. Do not write your name anywhere on these sheets. You will only be identified by your Laboratory and grade or subject-speciality, so please be realistic in your record.
2. Begin the record immediately—subject only to sickness or leave. You may think that just now is a bad time and will yield a non-typical result. This may be true of you, but the consolidated R. & D. answers are likely to be quite typical.
3. When you start, enter Date in the top right-hand space on each sheet, together with the date of thirteen days later. Once started, do not try to "compensate" for days when no reading is done, by recording extra days outside the period.
4. Keep the Diary with you (including off duty) except when you know you will not be doing any scientific or technical reading.
5. Confine your record to the kinds of literature shown: for practical reasons we have omitted data tables and handbooks, dictionaries, patents, standards, etc. If you fill a sheet, please continue in the spare book herewith.
6. When your Diary is completed, send the whole book (and the spare one) direct to :—

J. Roland Smith, Library, I. G. H. Q., Risley.

SERIAL NO........ DATES .... to ....

Record of **ABSTRACT JOURNALS, Etc.**

read during the continuous fourteen-day period shown.

Instructions: Please record on this sheet (or a continuation sheet bearing the same dates) each time you use a published abstract journal, entering against its title a stroke / when you are using it to locate or identify past literature, or a "X" when you are using it to keep abreast of recent developments. If you use an abstract journal not on this list, please write-in its title at the end.

British ceramic abstracts	
British Non-Ferrous Metals Research Association Bulletin	
British Scientific Instrument Research Association Bulletin	
Chemical abstracts	
Corrosion (abstracts)	
Electrical engineering abstracts (Science abstracts B) ...	
Electronic and radio engineer (abstracts)	
Engineering index	
Engineer's digest	
Index aeronauticus	
Iron and Steel Institute Journal (abstracts)	
Journal of applied chemistry (abstracts)	

(continued overleaf)

ABSTRACT JOURNALS

Light metals bulletin	
Metallurgical abstracts	
Nickel bulletin	
Non-destructive testing (abstracts)	
Nuclear science abstracts	
P.E.R.A. abstract bulletin	
Physics abstracts (Science abstracts A)	
Rubber abstracts	
Vacuum (abstracts)	

SERIAL NO........

DATES ....
to ....

Record of PERIODICALS

(JOURNALS, REVIEW JOURNALS, PROCEEDINGS, TRANSACTIONS OF SOCIETIES) Etc.

read during the continuous fourteen-day period shown.

Instructions : Please record on this sheet (or a continuation sheet bearing the same dates) each separate Periodical, and each article in it which you either scan-through or read. For a periodical (and also for a photocopy or reprint from it) put a stroke/ in Col. A—the position does not matter. For each article consulted (and again for each photocopy or reprint of an article) put a stroke in each of Cols. B-F, choosing the sub-column which best describes the circumstances.

Avoid marking more than one sub-column in any main Col. for the same article. Start a fresh line for each article.

COL. A PERIODICALS, PHOTOCOPIES, REPRINTS	COL. B ARTICLES, PHOTOCOPIES, REPRINTS		COL. C WHERE READ					COL. D WHERE PERIODICAL (PHOTO., REPRINT) OBTAINED						COL. E HOW REFERENCE TO IT WAS FOUND							COL. F USE MADE OF ARTICLE (PHOTO., REPRINT)						
	(either) SCANNED	(or) READ	1 In the library.	2 In the office.	3 In the laboratory.	4 Travelling on duty.	5 Off duty (at home, etc)	6 Routine circulation by your library	7 Passed on by senior or colleague for information.	8 Requested from senior or colleague on loan	9 Your private subscription or free list copy.	10 From the library.	11 From other source.	12 Recommended by author, senior or colleague.	13 Your "hunch," memory, etc. or seen by chance in actual periodical.	14 Seen in "Information Bulletin" or subject-bibliography.	15 Personally found in published abstract-journal (e.g. Chem. Abs.).	16 Seen in previous or other periodical review, book.	17 Provided by Librarian or Information Officer.	18 No reference involved cf. 6, 7, 9.	19 For writing a Report, TN. etc.	20 For writing article, book, lecture, etc.	21 Information provided for a colleague or junior.	22 For your current or future research commitments.	23 For general interest.	24 Discarded—not of use.	25 Discarded—seen too late to be useful.

SERIAL NO........

DATES ....
to ....

Record of **REPORTS Etc. AND COMMITTEE PAPERS (NOT MINUTES)**

read during the continuous fourteen-day period shown.

Instructions: Please record on this sheet (or a continuation sheet bearing the same dates) each separate document you scan-through or read. Put a stroke / in each of Cols. A-E for each document, choosing the sub-column which best describes the circumstances. Avoid marking more than one sub-column in any main Col. for the same document. Start a fresh line for each document.

COL. A		COL. B WHERE READ					COL. C WHERE DOCUMENT WAS OBTAINED					COL. D HOW REFERENCE TO IT WAS FOUND							COL. E USE MADE OF DOCUMENT						
(either) SCANNED	(or) READ	1 In the library	2 In the office	3 In the laboratory	4 Travelling on duty	5 Off duty (at home, etc.)	6 Your name on Distribution List.	7 Passed on by senior or colleague for information.	8 Requested from senior or colleague on loan	9 From the library.	10 From other source.	11 Recommended by author, senior or colleague.	12 Your "hunch" memory, etc. or report seen by chance.	13 Seen in "Reports List" or subject bibliography.	14 Personally found in published abstract journal (UNCLASSIFIED only).	15 Seen in another report, etc.	16 Provided by Librarian or Information Officer.	17 No reference involved, cf. 6, 7	18 For writing a Report, TN, etc.	19 For writing article, book, lecture, etc.	20 Information provided for a colleague or junior.	21 For your current or future research commitments.	22 For general interest.	23 Discarded—not of use.	24 Discarded—seen too late to be useful.

SERIAL NO........

DATES ....
to ....

Record of **TEXTBOOKS, SYMPOSIA AND ANNUAL REVIEWS** (*e.g. Progress in Metal Physics*).

read during the continuous fourteen-day period shown.

Instructions: Please record on this sheet (or a continuation sheet bearing the same dates) each different volume you read or briefly consult. Put a stroke / in each of Cols. A-E for each volume, choosing the sub-column which best describes the circumstances.

Avoid marking more than one sub-column in any main Col. for the same volume. If the book is consulted several times in one day, record it only once; but make a new entry if you use it again next day. Start a fresh line for each volume.

COL. A		COL. B WHERE READ					COL. C WHERE BOOK WAS OBTAINED					COL. D HOW REFERENCE TO IT WAS FOUND								COL. E USE MADE OF MATTER READ OR CONSULTED						
(either) BRIEFLY CONSULTED	(or) CONTINUOUS READING	1 In the library.	2 In the office.	3 In the laboratory.	4 Travelling on duty.	5 Off duty (at home, etc.)	6 Passed on by senior or colleague for information.	7 Requested from senior or colleague, on loan.	8 Your own property.	9 From the library (including permanent loan)	10 From other source.	11 Recommended by author, senior or colleague.	12 Your "hunch" memory, etc. or book seen by chance.	13 Seen in "Book list" or subject-bibliography.	14 Personally found in published abstract-journal (e.g. Chem. Abs)	15 Personally found in library card catalogue.	16 Seen in other book, periodical, review, etc.	17 Provided by Librarian or Information Officer.	18 No reference involved, cf 6, 8.	19 For writing a Report, TN, etc.	20 For writing article, book, lecture, etc.	21 Information provided for a colleague or junior.	22 For your current or future research commitments.	23 For general interest.	24 Discarded—not of use.	25 Discarded—seen too late to be useful.

Methods by Which Research Workers Find Information

R.M.FISHENDEN

ABSTRACT. A survey has been made at the Atomic Energy Research Establishment, Harwell, to discover the methods by which research workers obtain the information they use and read. The object of the survey was to find which methods were most effective in bringing information to their notice, and so to improve the information services in the establishment. The survey was made by two methods: diary cards and personal interview.

The results showed that the following were the principal ways by which information was found. The figures represent percentages of *all items recorded* in the diary survey: regular reading of the current literature, including new reports, 29%; papers found through references in other papers, 9%; personal recommendation, 11%; and scanning lists of titles included in the report lists and information bulletin issued by library, 17%; *Nuclear Science Abstracts,* 7%; found for readers by the Library, 4%.

For the retrieval of old information (22% of all items recorded) there was a marked reliance on personal indexes (4%) and "previous use" (i.e., memory) (10%). All other retrieval methods combined accounted for only 8% of items. There is a strong inference that inadequate attention is paid to systematic searching of the literature and that greater use could be made of library services for such searches.

The use of the foreign language literature was small (5%) as was the use of reviews (4%).

Comparison with other records, comparison between the diary and interview surveys, and the general consistency of the figures indicates that the results of the diary survey were unexpectedly reliable. An important conclusion is that useful results can be obtained from a much simpler diary card than those used in some previous investigations.

The detailed results, relating as they do to a particular set of circumstances, are of limited general interest, but they give valuable and much needed guidance on the ways in which the AERE information services should be developed.

R.M.FISHENDEN Atomic Energy Research Establishment, Harwell, England.

1. SUMMARY OF SERVICES GIVEN BY A.E.R.E. LIBRARY

The Atomic Energy Research Establishment at Harwell is the Headquarters of the Research Group of the United Kingdom Atomic Energy Authority. It has a scientific and technical staff of nearly 2000, with several thousand ancillary staff. The Library serves the whole Establishment but for the purposes of this paper only the scientific and technical staff will be considered. The Establishment is divided into fourteen subject Divisions under Division Heads, e.g., Chemical Engineering, Chemistry, Engineering, Isotope, and Reactor, and each Division is broken down into Groups under Group leaders.

The library contains some 15,000 books, 12,000 volumes of periodicals, 22,000 pamphlets and 145,000 reports, and receives about 750 current periodicals. It is divided into three sections: *(a)* the Reading Room (published literature), *(b)* the classified and unclassified Report Rooms, and *(c)* the Information Office. The Reading and Report rooms are under qualified librarians, the Information Office is staffed by scientists and the whole is under a librarian with scientific qualifications.

1.1. READING ROOM AND REPORT ROOMS

These rooms provide reference and lending services and a photocopying service. They maintain the following catalogues: *(a)* Author catalogues of books, pamphlets, periodicals, U.K.A.E.A. reports and other reports which do not appear in *Nuclear Science Abstracts.* (N.S.A. is an abstract journal issued by the U.S. Atomic Energy Commission and is used at Harwell as a catalogue and accessions list of U.S.A.E.C. unclassified reports).

(b) Subject catalogue of books, pamphlets, and periodicals. It is classified by the U.D.C.

(c) Alphabetical subject catalogue of U.S.A.E.C. unclassified reports. This is provided by the U.S.A.E.C. The subject catalogue of all other reports is held by the Information Office.

(d) Serial catalogue (i.e., catalogues under series numbers) of classified and unclassified reports.

Accessions lists are issued: (1) recent additions (books, pamphlets, periodicals), monthly; (2) “Unclassified” and “Official Use Only” reports, weekly; (3) “Confidential” reports, fortnightly; (4) “Secret” reports, monthly.

U.S.A.E.C. unclassified reports appearing in *Nuclear Science Abstracts* (N.S.A.) are not listed. N.S.A. is widely distributed about the Establishment.

1.2. THE INFORMATION OFFICE

This office provides scientific and technical information and prepares bibliographies in answer to requests. It abstracts and classifies by U.D.C. the reports received in the library, other than those in N.S.A.; the abstract cards forming the subject catalogue to the report literature. Other services provided are:

(a) A weekly "Information Bulletin" of articles of interest in the current periodicals. Abstracts of the articles are not included. The list is intended as a quick notification of what is available.
(b) Notifications to individuals of articles in the current literature of interest to them (notification slips).
(c) Despatch of new reports to persons most interested in them.
(d) Translation from all languages except French. It is expected that members of the Establishment can read French.

1.3. DIVISIONAL LIBRARIES

These supplement the Main Library. Each Division has at least one. They are not subject to the control of the Librarian, but there is close liaison. Most of the material held by Divisional Libraries is also in the Main Library.

In addition to the above collections, authors of reports send copies to individuals on the Site by putting them on the initial distribution of the reports concerned.

2. CHOICE OF SURVEY METHOD

Two methods were considered for this survey: first, interviews of research staff by a member of the Information Office and second, the diary card method. The expected advantages and disadvantages of the two methods are summarised below.

2.1. DIARY METHOD

The advantage of the diary is that it offers, at least in theory, a method of securing an accurate record of the information used by the diary keeper and how the information was obtained. On the other hand, the method is more dependent than is the interview method on cooperation from diary keepers. Furthermore, the record is unlikely to be 100% complete, and omissions may not be random.

2.2. INTERVIEW METHOD

This method is much the more flexible, and allows discussion of interesting points raised by those interviewed. The main objections are that it is subjective, and that, since it relies on memory, recent events are likely to receive too much prominence.

In discussion prior to the survey, opposing views were strongly held. Critics of the interview method held that so much subjective information was already available on information methods that a more exact approach was required, while critics of the diary method held that the filling in of cards would be so erratic that any appearance of numerical accuracy would be illusory. So firmly were these views held that compromise proved impossible, and the decision was taken to conduct independent surveys by the two methods. This had the incidental advantage that the surveys could be planned so as to make some of the results comparable.

In the event the results of the two surveys were in fair agreement (see section 6), and the weaknesses of neither method seem to have been as serious as expected by its critics.

3. THE DIARY CARD INVESTIGATION

The diary card investigation of the methods used by scientists in finding information follows roughly on the lines of similar investigations by Bernal (1) and R.R.Shaw (2). However, a fundamental decision made at the start was that reliable results could be obtained only if record keeping was made very simple: the aim was a single mark on the card for each item recorded. The investigation also differed from those undertaken previously, in that it was carried out in a research establishment with a relatively large information office in the Library.

The diary card used, together with the supplementary notes given to diary keepers, are reproduced in Appendices 1 and 2. It will be seen that the card enables the diary keeper to show by a single dot how each useful “unit of information” was found, and whether it was in a report, review, book, or elsewhere. A second dot is required only for foreign language literature. Provision is also made to distinguish between background reading and information used directly on a job. It was expected that this information would be a small proportion of the whole, but that in getting it there would be a comparatively frequent use of some method of retrieval. If the figures were diluted with a larger number of background reading items (not involving retrieval problems), it was thought that it would be more difficult to identify the most effective retrieval methods. In the event, the distinction did not prove useful. [See *4.2 (b)*.]

Of the 17 horizontal lines on the card 1–10 cover information found by staff without library aids, and 11–17 information found with library aids.

The final layout of the card was decided after a one month pilot run in which members of the Chemistry Division at A.E.R.E. participated. This pilot run showed up a number of important omissions in the trial card, and also led to the attempt to separate “background” from immediately useful information.

The design of the diary card, and problems met in using it, are discussed further in Appendix 3.

The information on the completed cards was transferred to Hollerith punched cards, together with code numbers to provide the following information: *(a)* identity of diary keeper, *(b)* Division, *(c)* "pure" or "applied" research worker or engineer, *(d)* Group Leader or junior.

The survey was confined to staff in those grades containing only honours graduates. A sample of 80 staff was aimed at, and each Division was asked to provide volunteers pro rata with its size and to choose the volunteers so as to give a representative cross section of graduate staff. The cooperation from Divisions was sufficient to give a reasonably representative sample of the establishment, and 63 completed cards were returned. The diaries were kept for approximately 2 months.

4. RESULTS OF DIARY CARD SURVEY

The results of the diary card survey are presented in tables giving the following information: Table 1, overall results, i.e., totals for each box for all 63 participants; Table 2, number of participants making any use of particular methods; Tables 5 and 6, as for 1, but broken down into separate classes.

In addition, the record was discussed with 11 diary keepers. This is referred to subsequently as the "oral sample." There was surprising unanimity that for written information (columns A-H) the record was 80–90% complete and that in this field it presented a valid picture of the information used. All diary keepers found, however, that it was impracticable to find a logical basis for recording information received orally, or by written private communications.

4.1. OVERALL RESULTS

The overall results of the diary card survey are given in Table 1. Because the survey failed for oral information and written private communications, the results for columns I-P of the card have been excluded from the table (but see 4.2 *(k)* for figures on the use of Conference proceedings). The explanatory notes reproduced as Appendix 2 will be found helpful in interpreting the results.

The results in Table 1 are supplemented by Table 2, in which is shown the number of diary keepers making any positive use *during the survey period* of a particular service or information tool.

4.2. DISCUSSION OF OVERALL RESULTS

(a) Reading of the current literature is as expected the most important single method by which information is acquired. However, the results emphasize

TABLE 1 Overall results (63 diarists)

	KIND OF INFORMATION / HOW WAS INFORMATION FOUND	A	B	C	D	E	F	G	H	Q	R	TOTAL	%
		REPORT		PUBLISHED PAPER		REVIEW		BOOK		TRANSLATION	READ IN FOREIGN LANGUAGE		
		B/R	USE	B/R	USE	B/R	USE	B/R	USE				
1	REF. IN ANOTHER PAPER READ	9	28	32	70	5	10	3	7	2	14	164	9
2	REGULAR READING OF CURRENT LITERATURE	16	4	241	121	22	11	18	10	5	14	443	23
3	REPORT REC'D ON INITIAL DISTRIBUTION	43	61	3	12					1		119	6
4	PERSONAL RECOMMENDATION	37	58	52	42	2	7	7	10	1	6	215	11
5	NUCLEAR SCIENCE ABSTRACTS	62	37	10	14	1	1					125	7
6	CHEMICAL ABSTRACTS	1	3	15	14	1			1	3	3	35	2
7	PHYSICS ABSTRACTS	1	2	3	17	1				1	2	24	1
8	OTHER ABSTRACT JOURNAL	3	6	19	10				1	2		39	2
9	PRIVATE INDEX	10	23	4	29	1	1	4	11	2	3	83	4
10	PREVIOUS USE	3	50	4	38		5	21	65	7	3	186	10
11	I FOUND IT IN LIBRARY INDEXES	2	7	2	7		1	2	5	1		26	1
12	LIBRARY FOUND IT FOR ME	3	9	2	9			2				25	1
13	INFORMATION BULLETIN	7	4	83	35	3	3	2	2	3	7	139	7
14	LIBRARY REPORT LIST	107	56	11	6			1		9		181	10
15	BIBLIOGRAPHY PREPARED BY LIBRARY	2	1	1	4			3				11	1
16	LIBRARY NOTIFICATION SLIP	1	2	20	16					1		39	2
17	ISSUED ON LIBRARY INITIATIVE	27	9	3			1	1	1	3	7	42	2
	TOTAL	334	360	505	444	36	40	64	113	41	59	1896	
	%	18	19	27	23	2	2	3	6	(5%)			

that the current report literature is not read in the same way as current periodicals. Unless a reader is sent a personal copy of a report (row 3 of the diary card) he has to make his own selection of titles, and he does this principally from *Nuclear Science Abstracts* (row 5) and the report lists issued by the library (row 14).

TABLE 2 Proportion of diary keepers making any positive use of a particular source or kind of information

Nuclear Science Abstracts	32%
Chemical Abstracts	17%
Physics Abstracts	17%
Other abstract journals	19%
Private index	45%
Information bulletin	37%
Report lists	50%
Library notification slips	20%
Reviews	41%
Foreign literature	30%

(b) Although the number of items found from references in another paper is substantial, it was expected that this would be a relatively more important source. An independent survey which showed very limited cross-referencing between papers in a particular subject field, may provide the explanation. (It is hoped to report the results of this survey to the Conference.)

(c) Comparatively little use is made of abstract journals. Only one-third of the diary keepers made any positive (i.e., successful) use of *Nuclear Science Abstracts,* in spite of the fact that N.S.A. is widely distributed and is the only method of notifying staff of U.S. unclassified reports received (apart from the limited library service covered in rows 16 and 17). The sample oral enquiry confirms this conclusion.

(d) Twenty-eight (45%) of the diary keepers show some use of a private index, and it is doubtful whether the row 9 total reflects adequately their importance. Of the oral sample 8 had shown no use of a private index, but of these 5 maintained one and 2 said they had forgotten to enter its use in the diary. There is no doubt that these indexes represent a valuable source of information of which more use might be made if the owners were willing to receive other people's enquiries.

(e) Of the services offered by the Library, much the most widely used are the information bulletin and the report lists. 40% and 50% respectively of the diary keepers showed some positive use of these lists and the oral sample indicated that about half the remainder were looking at these lists, but had found nothing of interest during the period. Altogether an estimated 70% of diary keepers were using each list.

(f) The use made of other library services (rows 11, 12, 16, and 17) was

numerically comparatively small. This might be expected, as library indexes would only be used (or a request made to library staff) when other means had failed. The notification slips are intended to cover articles in journals which a worker would not be likely to see himself: a high proportion of the slips sent out were used (see Section 6). However there is a strong indication that inadequate use is made of library resources for literature searches (see *h*).

(g) Row 15 covers only the use of bibliographies already in existence. No bibliography was prepared for a diary keeper during the survey period.

(h) Most effective retrieval methods. Of the methods used for finding information, it is possible to make a rough division into those applicable to the current literature, and those applicable to retrieval of older material, as follows:

(i)	For current literature	Rows 2, 3, 5, 13, 14, 16, 17
(ii)	For retrieval	Rows 6, 7, 8, 9, 10, 11, 12, 15
(iii)	Not allocated	Rows 1, 4

Nuclear Science Abstracts is included in (i) since it is used principally as a list of the current U.S. report literature.

The numbers of items obtained by the different retrieval methods are shown separately in Table 3. It will be seen that two retrieval methods, private index and previous use, account for nearly two-thirds of the total. It seems a reasonable inference that the extent to which careful literature searches are made is small, and that greater use of library services for this purpose could profitably be made.

TABLE 3 Most effective retrieval methods

Retrieval method	*Items*	*%*
Abstract journals	98	23
Private index	83	19
Previous use	186	43
I found it in library indexes	26	6
Library found it for me	25	6
Bibliography	11	3
Total	429	100

Incidentally, the separation of the record into "background reading" and "use" items did not assist in identifying the most effective retrieval methods. However, as might be expected, three-quarters of the "retrieved" items were shown in the "use" columns and only one-fourth as background reading.

(i) The use made of the foreign language literature was slight, roughly 5% of items recorded being from this source. [This is the same result as obtained by Shaw (2)]. Less than half the diary keepers used the foreign language

literature at all. This is a not unexpected result, but it demonstrates once more how little foreign language information is used. It would be interesting to try and assess the cost of overlooking the remainder.

(j) Although the distinction between "background reading" and information required for use on the job is not precise, the results indicate that reading is spread roughly equally between the two. Exceptions are that of items read in current periodicals, about two-thirds are background, whereas books are read only one-third for background.

(k) The use made of reviews as a source of information was unexpectedly small, about 4% of items coming from this source. Less than half the diary keepers used reviews at all. An explanation may be that in a relatively new field only a small proportion of the literature is included in reviews, but the result does not seem to bear out the views expressed at the 1948 Royal Society Conference. The use of Conference Proceedings slightly exceeded the use of reviews. (These are the columns I/P results, not included in the tables.)

(l) 1896 items of information were recorded by the 63 diary keepers, an average of almost exactly 30 each. (This excludes 262 items in columns I-P of the card, which were excluded from the results for the reasons given in Section 4.1.) There was a considerable spread between a minimum of 3 and maximum of 103. The distribution of items recorded per diary keeper is shown in Table 4. It will be seen from the table that 5 diarists recorded over 70 items each (in the range 77–103). All the rest recorded less than 65. If the 5 high scorers are omitted, the average for the rest drops from 30 to 25.

TABLE 4 Spread of number of items recorded per diary keeper

Items recorded	*Diarists in range*
0–10	13
11–20	16
21–30	10
31–40	10
41–50	3
51–60	3
61–70	3
71–80	1
81–90	2
91–100	1
101–110	1
Total	63

If allowance is made for Christmas holidays, the average number of recorded reading acts per week is about 4.3, somewhat lower than Shaw's figure of 5.8 (2).

4.3. BREAKDOWN OF RESULTS FOR DIFFERENT TYPES OF DIARIST

As mentioned earlier, the Hollerith card records were coded so that results could be totalled separately for senior and junior staff, and for workers in pure and applied research. The results are shown for columns A-H only as there was no significant difference in the use of the foreign language literature.

In addition, the figures for "background reading" and "use" are combined as there is no significant difference in the proportion of each between the broken-down results and the overall results.

Comparison of senior and junior staff. The results for senior and junior staff are compared in Table 5. The similarity between the results for senior and junior staff are remarkable. In the four columns A/B - G/H the greatest difference is 2%, (and the similarity is maintained when the 8 columns A-H are added separately). There are slightly greater differences in the methods used to find information. Senior staff receive more reports on initial distribution (as would be expected) and rely more on references in other papers. Junior staff rely more on previous use and abstract journals, and are slightly more willing to use library indexes (catalogues).

The sample contained 23 senior staff of roughly group leader status and 40 junior staff, the average number of items recorded being 27 for senior staff and 32 for junior.

The consistency of these figures encourages the belief that some reliance can be placed on the results of the survey.

Comparison of results for pure and applied research workers. The results for pure and applied research workers are shown in Table 6.

As might be expected, there are considerable differences between them. In particular the former make practically no use of the report literature, over 70% of their reading being in journals, whereas the latter make roughly equal use of the report literature and journals. This difference is reflected in relative use made of different sources of information. The "pure" group make noticeably more use of references in other papers, but surprisingly little use of abstract journals. The "applied" group make greater use of library services.

The sample contained 19 pure research workers and 34 applied, the balance being engineers or not easily allocated to any group. The proportion of pure research workers exceeds the establishment average (about 20%) because of the inclusion of 6 staff of the Medical Research Council. The average number of items recorded was 25 for pure and 37 for applied research workers. It should be noted that the Medical Research Council staff have an independent Library of their own, and make very little use of the services covered by rows 11–17 of the diary card.

TABLE 5 Comparison of senior and junior staffs

	How was information found \ Kind of information	Senior staff (23 diarists)						Junior staff (40 diarists)					
		A/B Report	C/D Published paper	E/F Review	G/H Book	Total	%	A/B Report	C/D Published paper	E/F Review	G/H Book	Total	%
1	Ref. in another paper read	18	35	10	7	70	11	19	67	5	3	94	7
2	Regular reading of current literature	8	116	5	12	141	23	12	246	28	16	302	24
3	Report rec'd on initial distribution	45	3			48	8	59	12			71	6
4	Personal recommendation	33	37	1	2	73	12	62	57	8	15	142	11
5	Nuclear Science Abstracts	25	4			29	5	74	20	2		96	7
6	Chemical Abstracts	3	7	1		11	2	1	22		1	24	2
7	Physics Abstracts	1	10	1		12	2	2	10			12	1
8	Other abstract journal	2	6		1	9	1	7	23			30	2
9	Private index	5	10	1	8	24	4	28	23	1	7	59	5
10	Previous use	2	5		29	36	6	51	37	5	57	150	12
11	I found it in library indexes	4				4	1	5	9	1	7	22	2
12	Library found it for me	7	4			11	2	5	7		2	14	1
13	Information bulletin	6	33		1	40	6	5	85	6	3	99	8
14	Library report list	58	10			68	11	105	7		1	113	9
15	Bibliography prepared by library		2		1	3	-	3	3		2	8	1
16	Library notification slip	2	16			18	3	1	20			21	2
17	Issued on library initiative	17	1	1		19	3	19	2		2	23	2
	Total	236	299	20	61	616		458	650	56	116	1280	
	%	38	48	3	10			36	51	4	9		

TABLE 6 Comparison of pure and applied research workers

		Pure research workers (19)						*Applied research workers (34)*					
	KIND OF INFORMATION / HOW WAS INFORMATION FOUND	A/B	C/D	E/F	G/H			A/B	C/D	E/F	G/H		
		REPORT	PUBLISHED PAPER	REVIEW	BOOK	TOTAL	%	REPORT	PUBLISHED PAPER	REVIEW	BOOK	TOTAL	%
1	REF. IN ANOTHER PAPER READ	5	58	13	8	84	17	31	43	2	1	77	6
2	REGULAR READING OF CURRENT LITERATURE	1	139	8	10	158	32	10	182	16	12	220	17
3	REPORT REC'D ON INITIAL DISTRIBUTION	10	1			11	2	72	14			86	7
4	PERSONAL RECOMMENDATION	8	56	2	5	71	15	80	36	7	11	134	11
5	NUCLEAR SCIENCE ABSTRACTS	3	3			6	1	96	19	2		117	9
6	CHEMICAL ABSTRACTS	1	7		1	9	2	3	22	1		26	2
7	PHYSICS ABSTRACTS		7	1		8	2	3	12			15	1
8	OTHER ABSTRACT JOURNAL		4			4	1	9	23		1	33	3
9	PRIVATE INDEX	1	8		1	10	2	29	22	1	6	58	5
10	PREVIOUS USE	8	16	4	21	49	10	44	25	1	61	131	10
11	I FOUND IT IN LIBRARY INDEXES		1		5	6	1	7	8	1	1	17	2
12	LIBRARY FOUND IT FOR ME		4			4	1	12	7		2	21	2
13	INFORMATION BULLETIN		27	1	1	29	6	11	87	2	2	102	8
14	LIBRARY REPORT LIST	11	5			16	3	144	9		1	154	12
15	BIBLIOGRAPHY PREPARED BY LIBRARY	1	5			6	1	2			3	5	
16	LIBRARY NOTIFICATION SLIP	1	9			10	2	2	27			29	2
17	ISSUED ON LIBRARY INITIATIVE	1	1			2		31	2	1	2	36	3
	TOTAL	51	351	29	52	483		586	538	34	103	1261	
	%	11	73	6	11			46	43	3	8		

5. THE INTERVIEW SURVEY

A representative sample of 50 graduate members of staff were selected at random for the interview survey. Each of the selected subjects was interviewed by Mr. B.Wilson, a member of the Information Office staff. The list of topics discussed at each interview is reproduced as Appendix 4.

The interview survey was designed primarily to obtain information of direct value to the Information Office in planning their services. Only those results bearing on the main theme of this paper are presented here, in the form of a comparison with corresponding results from the diary survey.

The results of Mr. Wilson's survey are being written up separately as an A.E.R.E. Library report and I have to thank him for permission to quote these extracts from them.

6. COMPARISON OF DIARY AND INTERVIEW SURVEYS

Both surveys give a figure for the proportion making any use of a particular reference source, and this has been chosen as a basis for comparison. The best comparison is possible in the case of the *Information Bulletin,* the *Reports* lists, and *Nuclear Science Abstracts.* For these the interviews gave figures for regular and occasional use. The diary gave figures for any positive use in the survey period and the oral sample indicated whether the remainder did not use the reference source, or had found nothing in it during the survey period. The figures, together with rather less complete figures for chemical abstracts and private indexes are included in Table 7.

TABLE 7 Proportion of sample making use of particular reference source

	Interview			*Diary*		
Source	*Regular use*	*Occasional use*	*Total*	*Positive use in survey period*	*Estimate of others who use*	*Approx. Total*
Information Bulletin	60%	6%	66%	37%	31%	70%
Report list	52%	6%	58%	50%	16%	65%
Nuclear Science Abstracts	40%	16%	56%	32%	17%	50%
Chemical Abstracts	22%	2%	24%	17%	—	—
Private Index	—	—	>60%	45%	30%	75%

The agreement between these figures is within the probable errors. A cruder comparison is possible between the results from the two surveys for the use made of bibliographies, of notification slips, and of reports issued on library initiative (rows 15, 16, and 17 of the diary card). Here, all that is available is a

"% using" figure from the interviews and a "positive use" figure from the diary cards.

	Interview "Any use" %	*Diary positive use*
Notification slips	54%	20%
Reports issued on library initiative	48%	24%
Library bibliographies	18%	12%

It was possible in the case of the notification slips to check the diary card results with library records. These showed that slips were sent to 13 of the diary keepers, and 13 showed a positive use of the service—a remarkable correlation. A total of 49 slips were sent out, and the diary records showed 39 items found from slips. This confirms that the records are reasonably complete, as recipients would not follow up all slips received.

The reason that the diary card figures are in each case lower than the interview figures is that they cover a shorter period. It is known that at least 44% of those interviewed were sent slips in the last few months, which is in reasonable agreement with the "any use" figure of 54%.

Although the interviews gave no figures comparable with the diary survey for the use of foreign literature, they confirmed that use was slight, and indicated in addition that important Russian and German material is not being used because staff do not request translations.

REFERENCES

1. BERNAL, J.D. The Royal Society Scientific Information Conference Report, Paper 46, 1948.
2. SHAW, R.R. Pilot study on the use of scientific literature by scientists. 1956.

APPENDIX 1

KIND OF INFORMATION → / HOW WAS INFORMATION FOUND ↓	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	
	REPORT		PUBLISHED PAPER		REVIEW		BOOK		WRITTEN P.C		ORAL P.C		LECTURE		CONF. PROCEEDINGS		TRANSLATION	READ IN FOREIGN LANGUAGE	
	B/R	USE	B/R	USE	B/R	USE	B/R	USE	B/R	USE	B/R	USE	B/R	USE	B/R	USE			
1 REF. IN ANOTHER PAPER READ																			1
2 REGULAR READING OF CURRENT LITERATURE																			2
3 REPORT REC'D ON INITIAL DISTRIBUTION																			3
4 PERSONAL RECOMMENDATION																			4
5 NUCLEAR SCIENCE ABSTRACTS																			5
6 CHEMICAL ABSTRACTS																			6
7 PHYSICS ABSTRACTS																			7
8 OTHER ABSTRACT JOURNAL																			8
9 PRIVATE INDEX																			9
10 PREVIOUS USE																			10
11 I FOUND IT IN LIBRARY INDEXES																			11
12 LIBRARY FOUND IT FOR ME																			12
13 INFORMATION BULLETIN																			13
14 LIBRARY REPORT LIST																			14
15 BIBLIOGRAPHY PREPARED BY LIBRARY																			15
16 LIBRARY NOTIFICATION SLIP																			16
17 ISSUED ON LIBRARY INITIATIVE																			17
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	

Instructions for use of Card

Make a single mark (preferably a dot) on the card for each useful item of information found (See separate explanatory notes for a definition of "item of information" and for more detailed guidance on use of the card). For the type of information acquired in background reading, use the columns headed "B/R". For information directly applicable to your job, use the columns headed "use". E.g. for a report found in N.S.A. of immediate use for your job put a dot in box B5.

For foreign language publications make an addition mark in column Q or R as appropriate.

Please complete the table below, but leave the right hand column blank.

		Hollerith coding
NAME		
GROUP		
DIVISION		
GRADE		
Special Exercise*		

Date of start of record .... Date of end of record ....

*Indicate here if this card was used to cover special literature survey prior to starting new job or writing report.

HD 3421

APPENDIX 2 NOTES ON USE OF DIARY CARDS

1. Make a single mark (preferably a dot) on the card for each useful “unit of information” obtained. The card is laid out so that you can indicate by this dot how you found the information (e.g. from an abstract journal, by receiving it on initial distribution, or by personal recommendation), what the information was in (e.g., a report, a journal, a review) and whether the information was for background reading (B/R) or direct application to a job (use). E.g., for an article found in the information bulletin of direct application to your job, put a dot in box D13.
2. “Unit of information” should normally be interpreted as a report, article, review, lecture, but make a record even if only part of the information in a report etc. is used.
3. For foreign publications make an *additional* dot in column Q or R as appropriate.
4. It is the intention that important information obtained in discussions (i.e., oral private communications) should be recorded. Boxes K4 and L4 will normally be appropriate.
5. If information reaches you several ways enter the first only.
6. If you come across information of a kind or by a method not provided for on the form, e.g., newspaper article, or pamphlet, don't record it. A pilot run with spare boxes for such items showed they were insignificant in number.
7. For borderline items between “background” and “direct application” items, make a dot on the borderline between the two relevant boxes.
8. There may sometimes be two stages in finding information, e.g. personal recommendation (have you seen that Brookhaven report by Smith and Brown) followed by identification in an index. If so, put a *cross* in both boxes.
9. If during the period you are keeping this diary you undertake a special literature survey e.g., because you are starting a new project or writing a report, please use a separate diary card for this.
10. Disregard the numbers written sideways. These are for use when results are transferred to Hollerith cards.
11. The validity of your record will not be affected by an occasional omission (provided these omissions are random!), so please don't give up because you forget to use the card for a few days.
12. If one of the boxes on your card gets full, start another card. The Divisional representative who issued your card has a stock of spares.

R.M.Fishenden

12th November, 1957.

APPENDIX 3 DISCUSSION OF DIARY CARD LAYOUT

In general the card layout appears to have been quite satisfactory, but the following points may be of interest to anyone attempting a similar survey:

(a) "Unit of information" is rather a vague term on which to base an investigation. Nevertheless, it is difficult to improve on. It appears to have been a satisfactory definition for written information. Only 2 people said the definition was so vague they could not keep a satisfactory diary, although a few others may have failed to keep a record for the same reason.

(b) The diary card failed to offer a practicable method of recording information received by the spoken word. No one could find a logical basis on which to decide what to record.

(c) The card makes no provision for recording useless information received. One or two people mentioned this, but it is doubtful whether the extra complication would be worth while.

(d) The distinction between "background reading" and "use" did not assist very much in identifying the most important retrieval tools. A better alternative might be to distinguish between current literature and "old" information.

(e) The card did not distinguish between information obtained (i) by attending lectures or conferences and (ii) by reading a written record of the proceedings.

APPENDIX 4 TOPICS DISCUSSED AT INTERVIEWS

1. Rank, division, position, length of service at Harwell.
2. Extent of use of *(a)* Library services and *(b)* Information Office services.
3. Proportion of use devoted to background reading and to specific searching.
4. Frequency of use of the following, and comments on their usefulness: *Information Bulletin;* accessions Lists (including reports and books); bibliographies; notification slips from daily scrutiny of journals; initial circulation of reports by the Information Office.
5. Frequency of consultation of the following: journals; books; reports; *Nuclear Science Abstracts;* other abstract journals; U.S.A.E.C.'s subject catalogue of unclassified reports; library author catalogue; Information Office (security classified); subject catalogue of reports; library subject catalogue; press cuttings; lantern slide catalogues; Microcards.
6. Extent of use of *(a)* Library staff, and *(b)* Information Office staff.
7. Personal indexes.
8. Use of foreign language material.
9. Journals on subscription and journals read or scanned regularly.

Questions were also asked on a number of topics of domestic interest to the Library, e.g., on the efficiency of certain services.

Determining Requirements for Atomic Energy Information from Reference Questions[1]

SAUL HERNER and MARY HERNER

In August of 1956, under a grant from the National Science Foundation, the writers undertook a study to determine the relative efficacy of tailor-made classification systems as a basis for coding and organizing scientific information by manual and mechanical means. One phase of this study involved the design of a special classification scheme for libraries in the atomic energy field. As a basis for developing logical categories by which the literature of atomic energy and related subjects might be broken down, a representative number of typical instances in which actual users of this literature asked questions of it were collected. These typical instances took the form of approximately 5000 reasonably current reference questions received by 14 atomic energy research and reference organizations in the United States. In addition to furnishing some insight into how typical users categorize the information contained (or which they hope is contained) in atomic energy literature, the analysis of reference questions was used to determine the working language and idiom of these users.[2]

At a certain point in the course of the analysis of the reference questions, it became clear that they might also be used to define the information requirements of users of atomic energy information. Obviously, a reference question is an expression of a need for information. If a statistical quantity of such expressions of need could be collected and categorized, it would seem to follow that they could furnish a basis, or at least a partial basis, for defining a group's

SAUL HERNER and MARY HERNER Herner and Company, Washington, D.C.

[1] The work discussed in the present paper was supported in part by a grant from the National Science Foundation.

[2] A paper by Fred R.Whaley, "Retrieval Questions from the Use of Linde's Indexing and Retrieval System," which has been placed in Area 4, also contains an analysis of reference questions.

information requirements, provided that they were collected in such a way as to characterize the intellectual level and subject interests of the entire spectrum of users. On the basis of the foregoing logic, a small tangent was taken from the study of classification systems, and an auxiliary study was made of the reference questions for their own sake.

The auxiliary study was undertaken with the understanding that the method being used has its limitations and pitfalls, as do all other methods for analyzing and defining people's problems. Reference questions received by libraries and similar reference agencies do not encompass all the problems within organizations or among groups of workers which require searches for information. It is probable that the majority of problems requiring access to information which does not exist immediately within one's own mind are solved either with no recourse whatever to formal tools such as libraries, or by using the publications in libraries, with no direct reference assistance from librarians.

Thus, a study of reference questions, such as the present one, is actually a study of a certain, narrow type of information requirement: one which leads or permits the information seeker, for one reason or other, to place the task of getting the information he needs in the hands of a person or group outside of himself. But, whatever the motive behind reference questions, their study on a quantitative and qualitative basis would seem to be a means of defining those information requirements which reference librarians and other information specialists are likely to be called upon to meet. The problem is to collect such questions in sufficient quantities as to be statistically significant and representative.

METHOD OF STUDY

The reference questions in the present study were collected over a one-year period, beginning in the fall of 1956 and ending in the fall of 1957. The method of collection was as follows. Through the medium of a meeting conducted by the Technical Information Service of the United States Atomic Energy Commission, the Commission's national laboratories and other prime contractors were asked to collect their current reference questions and to forward them to the authors' firm. In order to facilitate the transcription and forwarding of questions, each of the participating organizations was given a supply of specially designed forms on which they were to fill in, for each question, the name of the organization receiving the question and the question itself, exactly as it was received. The participating organizations were asked to forward *all* questions received, regardless of whether they were technical or non-technical.

The foregoing method of collection produced a total of 4696 questions from 14 institutions and organizations. In addition to Atomic Energy Commission prime contractors, the cooperating institutions and organizations included the U.S. Department of Commerce Office of Technical Services, which is a primary disseminator of atomic energy information to individuals and organizations outside the sphere of Atomic Energy Commission contractors. The full list of cooperating institutions and organizations is given in Appendix I.

At the end of the collection period, the questions were categorized first as to subject, second as to number of concepts, and third as to the logical relationship of the concepts to one another, in cases where questions involved two or more concepts.

Of the 4696 questions collected, 3851, or 82.0 per cent, were scientific or technical, involving one or more of the natural or engineering sciences, and 845, or 18.0 per cent, were non-technical. It has been suggested that nontechnical questions are often received on such a routine or casual basis that a significant proportion may have not been recorded and forwarded in the present study. Thus, the number of non-technical questions may in reality be somewhat higher than is indicated. However, the relative proportions of *kinds* of non-technical questions would appear to be valid, as would the relative proportions of the various kinds of technical questions.

THE NATURE OF THE REFERENCE QUESTIONS

The relative proportions of the various categories of non-technical questions are given in Appendix II. The most striking characteristic of these categories is the sharp falling-off after the first three. These, along with their relative percentages, were the following. Business and management techniques (31.3 per cent), Buyers' information and prices and business and commodity statistics (29.2 per cent), and information about institutions and organizations (16.4 per cent). These first three categories constituted 77 per cent of the non-technical questions, indicating the formidable role of business and related matters in atomic energy research and development programs.

The relative proportions of the various categories of scientific and technical questions are given in Appendix III. In the case of the scientific and technical questions, there was slightly more scatter among the categories, with the decrease from the highest proportions to the lowest a good deal more gradual than in the case of the non-technical questions. Nevertheless, the first three of the twelve categories of scientific and technical questions constituted two-thirds of the total. These first three categories, and their respective percentages of the total number of scientific and technical reference questions, are as follows: Description of a process or method of procedure (25.5 per cent), physical, chemical, and engineering properties of substances (24.6 per cent), and description of apparatus or equipment (16.8 per cent). In view of the period during which the questions were collected for the present study, the relatively small interest in radiation effects (2.9 per cent) is rather surprising. But this in itself is perhaps an indication of the advantages of quantitative measurement of requirements as opposed to the use of one's imagination.

THE CONCEPTUAL STRUCTURE OF THE REFERENCE QUESTIONS

Another means by which information requirements may be determined and expressed is in terms of the number of concepts contained in reference questions and how these concepts relate to one another logically. This type of analysis has significance primarily in cases where reference questions are put to storage and retrieval systems capable of performing correlations involving two or more concepts in their searches. As a rule, it is more difficult for a retrieval system, whether manual or mechanical, to perform a search on a subject involving a number of concepts than it is to perform one involving only one concept; and, as the numbers of concepts in questions increase, the searching difficulty increases.

The question of the numbers of concepts contained in reference questions, as well as the relationship of these concepts to one another, has been a subject of broad discussion recently, particularly as interest in mechanical storage and retrieval devices has heightened. It occurred to the authors that their accumulation of 3851 reasonably verbatim technical reference questions presented an interesting opportunity for measuring conceptual complexity on a statistically valid basis. In order to take advantage of this opportunity, the number of discrete concepts in each of the technical reference questions was counted and tabulated. For purposes of the count, a very strict definition was placed on “discrete concepts,” which were taken to mean those significant concepts in a question which could not be subdivided without changing their essential meanings. Thus, the question, “Give me information on engineering in nuclear reactors,” was taken to contain two concepts, “engineering” and “nuclear reactors,” and the question, “Has a hydrogen-fluorine torch been invented?” was considered to contain one concept, “hydrogen-fluorine torch.” In the case of the second question, the word “invented” was not counted as a concept because it was not considered significant to the meaning or understanding of the information that was wanted.

Using the foregoing definition and limitations, a count was made of the concepts contained in the 3851 questions. The results were as follows:

No. of Concepts	1	2	3	4	5	6
No. of questions	466	1818	1167	327	73	0
Per cent of questions	12.1	47.2	30.3	8.5	1.9	0

In general, the foregoing statistics are in agreement with the results of previous studies of this kind, with the bulk of reference questions containing three or less concepts. This apparent limit on the number of concepts likely to be encountered in reference questions is one which should be taken into account in the design of retrieval mechanisms and capabilities in information storage and retrieval systems.

Another question to be pondered in the design of retrieval systems, if they are to reflect actual user requirements, is that of the logical relationship of the concepts contained in a reference question. In order to determine the relationship of these concepts to one another, an analysis and count was made of the numbers of questions in which the concepts constituted logical sums (where the requestor would settle for information about concept A or concept B), or logical products (where the requestor had to have information about concept A *and* B), or logical differences (where the requestor was interested in concept A, but *not* concept B).

The results of these counts were as follows:

	Logical products	*Logical sums*	*Logical differences*	*Totals*
Number of questions (with more than one concept)	3773	45	33	3851
Per cent of questions (with more than one concept)	98.0	1.2	0.8	100

Here again, we have a demonstration of the usefulness of studies of actual demands made on information systems in determining how these systems should be set up. While much is made in the literature of the need for retrieval systems and devices that can handle logical sums, products, and differences, it develops, in the present case at least, that the vast majority of questions involve logical products and that questions involving logical sums or differences are relatively rare. Thus, a system that could only handle logical products would be entirely adequate in the present case.

It is, of course, entirely possible that a question may actually be a combination of a number of questions, each of which might involve logical sums, products, or differences. Such questions are frequently discussed by writers in storage and retrieval systems. However, in the present case, from the count that was made of the concepts in the questions studied, it would appear that reference questions involving such complexities are more likely to be more in the realm of hypothesis than reality. Here again, the advantages of analyzing user requirements before designing a storage and retrieval system are demonstrated.

CONCLUSIONS

From the foregoing discussion, it is evident that useful data on the information requirements of a body of users can be obtained from collecting and analyzing statistical quantities of their reference questions. It is entirely probable that the results of a study such as the present one, based on questions from workers in a field other than atomic energy, would produce results quite different from those obtained in the present case. It is also possible that a study of reference questions in the atomic energy field, conducted at some future time, would also produce results at variance with the present ones, first, because the field and the interests of workers in the field are bound to change, and, second, because the capabilities of storage and retrieval mechanisms are likely to improve and offer the searcher greater opportunities for reference help.

APPENDIX I. ORGANIZATIONS AND INSTITUTIONS COOPERATING IN THE COLLECTION OF REFERENCE QUESTIONS

Oak Ridge National Laboratory Library, Oak Ridge, Tennessee

Sandia Corporation, Albuquerque, New Mexico

National Reactor Testing Station, Phillips Petroleum Company, Idaho Falls, Idaho

Atomics International, Canoga Park, California

National Lead Company of Ohio, Cincinnati, Ohio

Technical Library, Atomic Energy Commission, Washington, D.C.

University of California (Berkeley) Radiation Laboratory

Union Carbide Nuclear Company, Oak Ridge Gaseous Diffusion Plant Library, Oak Ridge, Tennessee

Technical Information Service Extension, Atomic Energy Commission, Oak Ridge, Tennessee

Brookhaven National Laboratory, Upton, New York

U.S. Department of Commerce Office of Technical Services

University of California (Los Angeles) Atomic Energy Project

General Electric Company, Hanford Laboratories Operation, Richland, Washington

E.I. du Pont de Nemours and Company, Explosives Department, Atomic Energy Division, Aiken, South Carolina

APPENDIX II. CATEGORIES OF NON-TECHNICAL QUESTIONS COLLECTED

	No. of questions	*Per cent*
Business and management techniques	265	31.3
Buyers' information and prices: business and commodity statistics	247	29.2
Information about institutions and organizations	139	16.4
Documentation and communication techniques	28	3.3
Spelling: non-technical definitions: identification of quotations	24	2.8
Meeting programs	22	2.6
Popular information on atomic energy	21	2.5
Safety statistics: general safety programs	20	2.4
Education and training	16	1.9
Laws and regulations	15	1.8
History: dates	14	1.7
Requests for bibliography of specific author	13	1.5
Geographical information	12	1.4
Biographical data	9	1.1
Totals	845	99.9

APPENDIX III. CATEGORIES OF TECHNICAL QUESTIONS COLLECTED

	No. of questions	*Per cent*
Description of a process or method of procedure	969	25.5
Physical, chemical, and engineering properties of substances	953	24.6
Description of apparatus or equipment	651	16.8
Physical and chemical constants	635	16.4
Biological effects of substances: Hazards: Toxicology	225	5.8
Radiation effects	112	2.9
Materials for specific applications	101	2.6
Composition of materials	54	1.4
Standards and specifications	46	1.2
Technical definitions	46	1.2
Description of meteorological or geological phenomena	39	1.0
Mathematical constants and methods	20	0.5
Totals	3851	99.9

Systematically Ascertaining Requirements of Scientists for Information

JIŘÍ SPIRIT and LADISLAV KOFNOVEC

At present, the increasing amount of scientific and technical literature renders it impossible for scientists and technical workers to follow up all the new literature related to their fields. In order to identify and obtain the most important documents, they inevitably need the assistance of information unit officers. These are in a position systematically to provide information about newly published documents by means of (abstracts) cards.[1]

If an information unit is to work for only a small number of research and technical workers, the information officers will be able to store up in their memories the different items that may be of interest to individual scientists. If, however, the number of specialized scientists or research workers is large, the information unit must be provided with special accessories if a systematic information service is to be correctly performed. In what follows, one of the most important of these accessories, i.e., the *thematical plan* of the establishment or institution, will be described in detail.

Obviously, a scientist cannot wait until some report of some foreign experience happens to come to his knowledge. Therefore, the information officer must know in advance what the individual creative workers will need. He must also know how and where the necessary information can be acquired.

In order to survey the requirements of the individual scientific workers, information officers must constantly be in contact with the scientists and choose useful information on the basis of the thematical plan. That is to say, each

JIŘÍ SPIRIT and LADISLAV KOFNOVEC Research Institute for Materials and Technology, Prague, Czechoslovakia.

[1] The paper by W.K.Lowry and J.C.Albrecht entitled "A Proposed Information Handling System for a Large Research Organization" (Area 5 of the Conference) also contains a discussion of methods for determining the information requirements of individual scientists served by an information service.

year, prior to the beginning of work on research, development, and design problems, the information officers will note, in a most detailed and specified way, the requirements of the individual scientific workers. These requirements should be arranged from two different points of view: (1) from a *personal* one, by making for each of the creative workers separately a specification of his requirements, classified, for example, according to the UDC; (2) from a *subject field* point of view, by systematically arranging the specified items according to the UDC, together with the name of the workers interested in them, all classified into larger groups for easy survey. These two specifications, if necessary supplemented by an alphabetical index of the main groups, represent a year's accessory tool for the information unit (or the technical library) and form the *thematical plan* which is the basis for providing foreign and domestic technical information.

The thematical plan is of the utmost importance in organizing a systematic information service, and is a significant component of any active information service. In order to gather the individual requirements of scientific and research workers, a special *“Information requirement sheet,”* of which a copy is appended, is filled in by the creative workers served by the information service.

On the information requirement sheet are shown the name of the scientific worker, his section and department, the number, designation and details of the research problem in question, and also the stage of its elaboration and the corresponding UDC number. It is most useful to indicate where best to search for information sources and during what period of time, and also to give characteristic details about the information required. Frequently a retrospective search is required, translations from various languages, etc. The back of the page provides space for entering data about accomplished searches or studies and translation work, and about the range of information services during the year. These “Information requirements sheets” are filed in the study department.

These “Information requirement sheets” are the basis for elaborating the establishment's thematical plan. Great care is taken in indicating UDC numbers for the different items since they are fundamental to the information unit's work for the coming year.

The first and basic part of the thematical plan is a systematic arrangement of all information requirements according to UDC, as for example,

	Steam boiler	
621.18	Steam boiler materials	Dr. Wild
621.18.004.6	Disturbances on steam power plants	Lindner
621.181.021	Steam boiler pipes	Dr. Wild
621.183	Steam boiler fittings	Dr. Wild
621.183.002.3	Materials for steam boiler fittings	dipl.Ing.Schmidt

621.186.1/.4	Steam line welding	Landgraf
621.187	Steam boiler operation	Dr. Wild

UDC numbers are indicated in the left column, the second column gives short descriptions of the required information, and the right column indicates the names of the scientific or research workers who wish to get information on the topics listed.

The second part of the thematical plan is arranged by departments of the establishment or by individual research groups of the Institute, with the names of the research workers in alphabetical order. Under the name of each research worker are all his information requirements and the respective UDC numbers, as for example,

Dipl.Ing.Schmidt

539.3.19 539.4.014 620.17 539.319 621.986	Methods for measuring residual stresses
	Stresses after cold forming
539.4.015 621.986	Influence of cold forming upon mechanical and physical properties of materials
620.18	Metallography, microscopy, apparatus for making polished sections
620.182	
620.183.002.3	Materials for steam boiler fittings

When abstracting a book or another nonperiodical publication or some journal literature, the information officer states, with the aid of UDC numbers, which of the scientific workers are to be supplied with the respective information. Thus, the systematization of the information service is assured, every research worker mentioned in the thematical plan receiving, for his personal use, copies of abstract cards, prepared in the study department, pertaining to his field of interest.

In establishments where the information service is already in common use, supplying copies of abstract cards, as mentioned, will be sufficient. In establishments where such services have not yet been introduced or are only occasionally used, the abstracts may be written on special forms with explanatory text such as: “We call your attention to the following article…. The periodical in question is at your disposal in our library,” etc.

The thematical plan is designed to be used not only by workers of the information unit, but also by scientific and technical workers, since it provides them with a good survey of the problems under study and of those who are working on them and who are in a position to give advice in certain cases. This is especially advantageous in large establishments or institutes where it is quite difficult to get a clear picture of the activities of the individual workers.

RESEARCH INSTITUTE FOR MATERIALS AND TECHNOLOGY
STUDY AND INFORMATION SERVICE

Information Requirement Sheet for 1958

UDC	*Problem number and designation:*	*Section: A*
539.56	Brittleness of some constructional materials	Dept.
539.4.012	especially as regards their durability and	Research worker:
539.419	overload capacity at alternating stress	Dipl.Ing. Muller

Details of the problem and its state: Problem to be handled from:
1.1.1957 to 15.2.1958

A complete information search to be elaborated prior to the beginning of study works on the research problem, i.e., until 31.12. 1956

Abstract cards are to be continually supplied later on.

Main and alternative information required

Influence of brittleness upon higher limits of fatigue at alternating stresses
Influence of overloading up to fatigue limit
Brittleness of constructional steels in general
Influence of grain size upon brittleness
Temper brittleness
Methods for brittleness testing
Notched-bar impact test
Tensile strength test

Information sources required:

Time limit: from 1949

Local limit: (only from German, Russian, etc.): from any available source

Characteristics of required information:

- ○ elementary theory (textbooks)
- • higher theories (monographs)
- • experimental methods
- • test results
- • practical operation
- • actual news

Kind of documents required:

- • books
- ○ articles
- • cumulative searches
- • translations
- ○ commercial literature
- ○ standards
- • study and research reports
- ○ other documents

Kind of requirements:

a. Cumulative search work (up to) 31.12.1956 finish 27.12.1956
b. Permanent positive search work—supply of abstract cards of actual news: yes
c. Translations (approx. number of typed pages):
from which language: Swedish

1. *a. Active research work:*

Cumulative search work (Nr R220)	finished on	27.12.1956
study work (Nr 12)	finished on	
	ordered on	

Cumulative search	• own ones	• Centr.Techn.Libr.	• Index of searches
Translations	• own ones	• Centr.Techn.Libr.	○ Index of translat.
Cumulative search card file	• own ones	• Centr.Techn.Libr.	○ other sources
Libraries	• own ones	○ Centr.Techn.Libr.	○ other sources
Standards	○ GOST ○ ASTM ○ BS ○ NF ○ TGL,DIN ○ other sources		

Technical reviews and bibliographies searched (title and year):

1. The *Engineering Index* 1949–1955
2.
3.
4.

Factories, institutes, universities, etc. visited:
Czechoslovak Technical Library

1. *b. Translations:*

Author	Title	Translation	Pages	Language	Ordered Finished	UTEIN[a] No.	Translated by

2. *Permanent search work—abstract cards*

Date	*Jan. 2*	*Jan. 7*	*Feb. 5*	*Feb. 18*	*March 1*
Abstract cards:	2	4	3	1	3
Own ones:	2	2	1	—	3
From external sources		2	2	1	—

[a] UTEIN = Central evidence of translations.

How Scientists Actually Learn of Work Important to Them

BENTLEY GLASS and SHARON H.NORWOOD

In the light of the very large sums of money—to say nothing of the time and skilled labor—expended annually on the indexing and abstracting of the scientific literature and on the development of new methods of recording and retrieving information, it seemed desirable to examine the actual ways in which representative scientists in practice find out about the existence of scientific work that is crucial to the development of their own research. The method we used was simple in the extreme. In individual interviews each scientist was asked to select a recent, significant paper from his own list of publications. From the references given in this paper he was then asked to choose up to five or six items representing scientific concepts and research of major or crucial significance to the development of his own work reported in the chosen paper. This done, the scientist was asked two questions: (1) How did you first learn of the existence of the work reported in each of the selected items? (2) Would it have made any significant difference to the progress of your own work had you learned of it sooner than you did? The first question was accompanied by an enumeration of the various ways in which experience soon showed the investigators had actually discovered work of importance to them (e.g., in casual conversation; from a formal report at a meeting; in a journal subscribed to by the investigator; in a journal regularly scanned in the library; in an abstracting service; in an indexing service; through a reprint received in exchange; from a reference book; from a review article on the subject; from a cross citation in some other article; in a formal discussion group; from a bibliography; or from a co-worker in the same laboratory.

The interviews elicited a vigorous response. Many of the investigators volunteered comments on the general problem of keeping up with the literature and on their personal methods of trying to cope with it. Fifty scientists were interviewed, representing a variety of fields, although concentrated in the biological

BENTLEY GLASS and SHARON H.NORWOOD Johns Hopkins University, Baltimore, Maryland.

sciences. There were 10 animal and human physiologists, 2 plant physiologists, 5 biophysicists, 8 biochemists, 2 organic chemists, 6 cytogeneticists and geneticists, 3 embryologists, 1 microbiologist, 1 physical anthropologist, 1 parasitologist, 6 psychologists, 1 oceanographer, 1 geologist, 2 sanitary engineers, and 1 radio engineer. The number of items from the literature on which they reported totaled 346. When the results were tabulated, rather striking differences were apparent between the workers in different fields. However, since the sample is too small for a really adequate representation of any single group, only the total for the entire sample will be presented. Even so, no positive conclusions should be drawn. The survey is presented merely as a pilot study in order to indicate a method for obtaining information about the actual methods employed by university scientists in their efforts to keep up with the progress of their own fields. The results, ranked in order of frequency, are given in Table 1.

TABLE 1 Ranking of Methods Whereby 50 Representative Scientists Actually Learned of Work Crucial to Their Own

	Number	*Percentage*
Casual conversation	78	22.6
From a journal regularly scanned	76	22.0
From a journal subscribed to	29	8.4
From a cross citation in another paper	24	6.9
Can't remember—general background—common knowledge	22	6.4
From a reprint received from the author	20	5.8
Through an abstracting service	18	5.2
From a co-worker in the same laboratory or department	15	4.3
From a reference work or textbook	15	4.3
From a review article (old work)	14	4.0
Through a formal report at a meeting	9	2.6
By chance	9	2.6
From a bibliography or material supplied in a course	6	1.7
Through an indexing service	4	1.2
In a formal discussion group	4	1.2
From a book list	3	0.9

It may be of some significance that the animal physiologists and biophysicists never referred to cross citations as valuable to them; the biochemists never reported discoveries in journals they regularly subscribe to (one wonders why?); the biophysicists, embryologists, and geneticists almost without exception eschew abstracting services; and the biophysicists and psychologists fail to use reprints.

Little weight can be placed upon the answers to the second question. However, 34 of the 50 scientists interviewed gave "No" as their answer for all their references, when asked whether it would have made a significant difference to the progress of their own work had they known of the related work sooner; and 9 answered "yes" at least once. A more extensive study of this relationship appears desirable.

The comments made by the scientists who were interviewed contrasted considerably with the data they provided. Many of them spoke of the importance of abstracts, the use they made of them, and the need for a better coverage of the literature and prompter publication. The data show that for this group of scientists abstracts were not often of primary importance. On the contrary, the most striking fact is unquestionably the very heavy reliance by most workers, and in nearly every field checked, on verbal communication with scientists working in the same area. It is not clear how this could apply to foreign work and workers, and it may therefore indicate a growing tendency to provincialism on the part of the scientists of the United States. In any future study of this sort, it would be valuable to have a breakdown of the distribution of items according to native country, foreign country of same language, and foreign country of different language; and to check this distribution against estimates of the amount of work being done in each field in the United States and abroad.

Dependence upon the memory of the scientists interviewed constitutes a flaw in the present procedure. In any future analysis of like nature the subjects might well be notified of the project in advance and asked to keep a record of how they learn about papers important to their own work, for the next 50 such items.

An extension of the study to include a larger and more representative group of American scientists would seem worthwhile, and if a group of European or other foreign scientists could be made the subjects of a similar survey, the international comparison would be very interesting.

Planned and Unplanned Scientific Communication

HERBERT MENZEL

Under a grant from the National Science Foundation of Washington, D.C., the Bureau of Applied Social Research of Columbia University has undertaken to explore ways in which communication research by interview survey methods can contribute to an understanding of the needs and means of scientific information-exchange. On the basis of such an understanding, proposals to improve scientific communication might be generated and evaluated. As a first step, it was decided to study the information-exchanging behavior of the biochemists, chemists, and zoologists on the faculty of a single academic institution—a prominent American university[1]. This paper reports selected results. A more complete account is on deposit with the National Science Foundation.

The objectives of the research that is ultimately envisaged had been defined as follows:

1. To distinguish the types of informational *needs* which scientists have, and to determine in what respects they remain unsatisfied.
2. To examine the *means* and occasions of scientific information-exchange,

HERBERT MENZEL Bureau of Applied Social Research, Columbia University, New York.

This paper is Publication A-259 of the Bureau of Applied Social Research, Columbia University. It is based on a pilot study carried out at the Bureau under the supervision of Charles Y.Glock. William A.Glaser and Robert H.Somers collaborated with the author in the execution of the work.

The Bureau operated under a grant from the Office of Scientific Information of the National Science Foundation. The encouragement given this work by Helen L.Brownson and Harry Alpert of the Foundation is gratefully acknowledged. Special thanks are due the biochemists, chemists, and zoologists whose generous contributions of interviewing time and attention made this work possible. Their visible interest in the matters discussed was a source of continuous stimulation.

[1] The 77 scientists whose interviews are analyzed here include all but 8 of the following: teaching faculty in biochemistry; teaching faculty and research associates in chemistry and zoology; provided they were in residence on the campus of the university during the spring of 1957. (Four biochemists, one chemist, and one zoologist refused to be interviewed, or to complete an interrupted interview. Two zoologists were interviewed for background information only.)

in order to single out the features which make them more or less able to meet the scientist's several needs.

3. To analyze characteristics of the scientist's specialty, his institution, and his outlook as possible *conditions* which influence his needs for information, his opportunities for satisfying them, and, hence, his information-gathering habits and felt satisfactions.

The exploratory study was intended to define problems, categories, and procedures for more systematic investigation. Although this report contains numerous frequency counts based on interview responses, they are to be regarded as illustrations of the possible outcome of further work and not as reliable findings. They may not even reliably describe the three academic departments studied, since the interview schedule was continuously modified and developed as the work proceeded.[2]

A SYSTEMIC VIEW

While the population of the initial study is small, it was decided to cast a broad conceptual net: to consider *all* the channels through which scientists exchange and gather information, and *all* the functions which scientific communication facilities are called upon to perform. In fact, the functions which the facilities must serve were made the organizing principle of the study. Rather than to ask, at the outset, “How well does this journal perform? How much does this meeting accomplish? What is wrong with that indexing system?” it was decided to ask as a first set of questions: “What are the functions of the scientific communication system? What mechanisms are now available for performing them? What are the inadequacies in the present performance of each function?”

This decision was founded on the belief that specific topics for investigation can be wisely selected and defined only after the broader context has been scanned. Furthermore, studies of communication processes among non-scientific publics[3] had shown that different communication functions are often performed, at their best, through different channels, and that the diverse channels may supplement one another in intricate ways.

Accordingly, we include in our purview not only the scientific literature and

[2] The schedule contained both structured and unstructured questions. A copy of the most recently used version is appended. The average interview took 1.9 hours.

[3] Many such studies are reviewed in Elihu Katz and Paul F.Lazarsfeld's *Personal Influence: The Part Played by People in the Flow of Mass Communication,* Glencoe, Ill.: The Free Press, 1955, and in Elihu Katz' The two-step flow of communication, *Public Opinion Quarterly, 21,* 1957, pp. 61– 78.

its manifold storing, abstracting, and indexing appendages; not only the formally established meetings and conferences, but also the informal, person-to-person modes of communication like correspondence, visits, and corridor conversations.

Secondly, we conceive of scientific communication as not necessarily limited to simple transactions between an individual scientist and a source of information. Communication includes more complex processes: several different channels of communication may have to interact to complete a transaction; one or more persons may serve as relays between the source of a message and its ultimate consumer; and contacts at each intervening step may be initiated now by the receiver, now by the bringer of the message. For these reasons we shall speak, somewhat loosely, of the "scientific communication system," meaning the totality of all publications, facilities, occasions, institutional arrangements, and customs which affect the direct or indirect transmission of scientific messages among scientists.

Thirdly, we believe that policies to improve the scientific communication system must be planned in terms of the entire range of its contributions to scientific progress, and not only in terms of the most obviously necessary informational services. At the present time, most plans are quite naturally directed at maximizing the efficiency of the system in the performance of its two most obvious functions: that of bringing scientists the available answers to specific questions, and that of keeping them abreast of current developments in given areas. Yet the policies that recommend themselves for these purposes may not be adequate to assure the fulfillment of other functions of the scientific communication system; and it is just possible that some of these same policies, say the shortening of papers at meetings or the streamlining of periodicals, may be detrimental to the system's other functions, which are not so obvious, but nevertheless important.

MULTIPLE FUNCTIONS OF THE SCIENTIFIC COMMUNICATION SYSTEM

The scientific communication system serves the progress of science not merely through the reference services it performs and through keeping scientists up to date in their chosen areas of attention. It serves in a variety of other ways as well: by enabling scientists to brush up on past work in additional areas; to verify the reliability of one source of information through the testimony of another; to ascertain the current demand for research on given topics; to locate rare materials; and so on. In fact it would be a mistake to think that the functions of scientific communication for the progress of science are limited to satisfying the informational needs of which each scientist is aware; that the only important job of the scientific communication system is, so to speak, to give each scientist what he wants, and knows he wants.

One important function of scientific communication which transcends the informational requirements each scientist can define for himself is that of directing the scientist's attention to new topics beyond those with which he has “kept up” in the past. Another is to assure the eliciting of suggestions and criticisms from fellow scientists. These and other rarely considered functions of the scientific communication system, and some of the mechanisms by which they are satisfied, are discussed in the report which has been deposited with the National Science Foundation.

KEEPING SCIENTISTS UP TO DATE

Yet even the performance of functions of which everyone is well aware requires more than the prompt appearance of information in the official channels (journals, meetings, etc.) and more than painless access to these media. It is, for example, not these formal media alone that keep scientists informed of current developments in their chosen areas of attention, in spite of the prodigious amounts of planned effort devoted to this communication function by individual scientists as well as by the professional organizations and publishers.

In fact, the news which comes to the attention of scientists is not restricted to the information obtained when they intentionally “gather information,” as it is called. Fortunately so! For a good deal of the news which comes to their attention in unplanned and unexpected ways, during activities undertaken and on occasions sought out for quite different purposes, proves to be of considerable significance to them. At least this was a frequent experience among the scientists we interviewed, in spite of the fact that their intentional activities for gathering information about current developments ranged all the way from the assiduous perusal of current periodicals to the button-holing of colleagues who had returned from conferences.

It was thought that it would be instructive to examine instances of significant scientific news coming to the attention of scientists through other ways than those which they systematically employ to “keep up.” This line of investigation was included in our study not only to learn about the operation of communication through informal and personal channels; by implication, it was thought, this approach would also throw light on possible inadequacies in the formally established methods of bringing current news to the scientist. Under an ideally functioning communication system, it was thought, the routinized and regular methods of gathering information would convey to each man all the scientific news that is pertinent to his work. Any pertinent information that actually reaches a man in some extra-routine or accidental way would therefore indicate a service not adequately performed through the routine system. As will be seen below, however, this form of interrogatory also taught some unanticipated lessons which are, perhaps, even more important.

In order to obtain records of instances of useful scientific news obtained during activities not undertaken for this purpose, the following question was asked:

Have there been any instances where some unlooked-for piece of information came your way that turned out to have bearing on your work? (If Yes) Tell me about the last time this happened.

Supplementary questions were asked in order to obtain complete accounts of the experiences. Not all the replies given proved pertinent to the present topic. Excluded from the list finally used were all accounts of information obtained in the course of routinely scanning the literature, attending meetings, or engaging in any other activity which was explicitly designed to find out what is new. Also excluded were episodes of information learned in the course of ordinary intercourse with departmental colleagues. Thirty-five usable accounts were obtained.

UNPLANNED MECHANISMS

The extra-routine mechanisms by which these messages reached the interviewed scientists are of four basic types:

1. The scientist *searches the literature for one particular item of information, and in the process stumbles across another* which proves useful to him. This, of course, is in addition to the countless times when a scientist comes upon some useful information which he had not anticipated in the course of his routine perusal of journals, or in the course of listening to the program of meetings which he regularly attends. What is meant here is rather a scientist searching the literature in order to find the answer to some specific question, and coming across pertinent information of another sort, information which he would probably not have seen had it not been for the accident of his search for the first topic. Thus a paleontologist reported:[4]

This morning my assistant wanted information on the geology of Southern Britain at a certain time. The same journal happened to contain information which will be interesting to our formal analysis problem…. (Do you think you would have seen this particular item otherwise?) It is doubtful.

[4] Unless otherwise indicated, matter set in smaller type is quoted from scientists' statements during the interviews. Matter in brackets [ ] paraphrases or supplements scientists' statements. Matter in parentheses ( ) quotes what the interviewer said to the scientist.

2. The second kind of situation which leads to the unexpected acquisition of information pertinent to one's work occurs when a scientist, in the course of contact for another purpose, *informs a colleague of his current work* or of some obstacle which preoccupies him at the moment, *and is rewarded with an item of information* that becomes important to his progress. A zoologist made an unexpected find in this way:

I learned in this way the whole technique for solving the problem of the recording technique for…. This had baffled scientists for twenty years. I went to a man at… Institute to buy some wire—Dr…., in the…Laboratory. He is an able and imaginative fellow. In conversation, we talked about my research and problems, and he dropped the hint which enabled me to solve the problem of the recording technique.

3. Sometimes a scientist hears about new developments from *a colleague* who *volunteers the information while* they are thrown *together for another* purpose. (We exclude here information gleaned during corridor conversations at scientific meetings, or on any other occasion attended for the explicit purpose of gathering news.) This may happen frequently during informal visits by one scientist to the laboratory of another. For example:

I went to the…Institute two months ago to give a talk. I stopped to see Dr. A. He is working in a different field. He found that a certain substance crystallized under certain conditions. We are interested in finding many different kinds of crystals. We will try using his experimental methods here. What Dr. A. found may not ever be published by him—it was a side effect, as far as he was concerned.

4. There is a fourth manner in which information of immediate relevance comes to the attention of scientists by what appears to be accidental routes. Frequently a colleague will deliberately seek out a scientist whom he knows to be interested in the matter, *in order to convey to him some information* that he happens to have heard. Thus, for example, a biochemist:

started a new project because I heard that someone in Germany had positive results in a related field. He published it one-half year later…. (How did you hear about the German scientist?) He had sent his unpublished results to another man in America who knew my interests and told me.

And a chemist gives a very detailed and instructive account of such an incident:

One of the problems in our work is to do a certain chemical separation. Recently a friend of mine had been in Europe. He met a young German who was developing a new technique. So we now try to apply this to our problems. Neither my friend nor I knew about the existence of this procedure before the encounter. The young German had invented this. My friend had not been looking for it—he was going through Europe, visiting labs and drinking beer with the people at the various labs. This technique was of no particular interest to my friend, but he knew it would interest me, and told me when he got back.

Table 1 shows how many accounts of experiences of each of these four kinds were reported by the interviewed scientists. It also shows that about half of the messages transmitted in these personal ways were actually in print at the time.

TABLE 1. Useful information obtained "accidentally":[a] how obtained

Manner in which message reached the interviewed scientist	*Number of messages*		
Total	*Published*	*Not published*	
Found in the literature while searching for another topic	4	4	—
Contributed by a fellow scientist upon being informed of colleague's current work	13	8	5
Spontaneously mentioned by a colleague while together for another purpose	4	—	4
Specifically addressed to the interviewed scientist by a colleague	9	2	7
Other, or not specified	5	2	3
Total	35	16	19

[a] Exclusive of information learned during ordinary intercourse with departmental colleagues, while scanning the literature, while attending meetings, or while engaging in any other activity explicitly designed to find out what is new.

CONTENT OF INFORMATION OBTAINED

What was the content of these messages which reached their consumers in such unexpected ways? Ten of the messages informed the scientist of new findings or principles (a biological mutant described, an archaeological find reported, a chemical reaction performed, etc.); eight informed him of the existence of new techniques, procedures, or apparatus; four furnished him details on the performance or adaptation of a technique; five told him who was doing work on a given topic or from whom a particular material could be obtained (Table 2).

TABLE 2. Useful information obtained "accidentally":[a] content

	Number of messages		
Content of message	*Total*	*Published*	*Not published*
New findings or principles	10	6	4
New procedure or apparatus	8	3	5
Details on procedure	4	2	2
Who does what; where to obtain material	5	1	4
Not indicated	8	4	4
Total	35	16	19

[a] Exclusive of information learned through ordinary intercourse with departmental colleagues, while scanning the literature, while attending meetings, or while engaging in any other activity explicitly designed to find out what is new.

The figures in Tables 1 and 2, as most of the figures throughout this report, are to be regarded as no more than suggestive and illustrative of possible findings of more systematic studies, as mentioned earlier. What they do suggest is that the apparently accidental obtaining of pertinent information plays a large role in the work of the interviewed researchers. Examples of it come up again and again in our interview protocols.

INDIVIDUAL ACCIDENT—AGGREGATE REGULARITY?

Why should this manner of learning of new developments be so prevalent? Part of the reason must be sought in the nature of specialization among basic researchers at the top level. They not only specialize to a high degree, but they also delineate their specialties in highly individual and original ways; often no more than a small handful will be specializing in precisely the same area. All the possible ways of classifying content cannot possibly be taken into account in the organization of journals, in the indexing and abstracting services, or even in the selection of titles for papers. Any given researcher is likely to find that the way of classifying reports which would be most relevant for his purposes has not been used. Within the confines of a narrow field, he attempts to scan everything that comes out; but beyond that he must depend largely on friends who work in the adjoining specialties, yet know what is of interest to him, to flag the pertinent material for him.

You see what happens around here [says a biochemist]. Everyone knows what problems you're working on. Whenever you come across something which might be of interest to another you make a note of it. This way the individual is able to be acquainted with a lot more than he would be if he didn't have the others on the look-out too.

If this is true, it becomes imperative to consider the information network as a system, and not merely as an aggregate of information-dispensing or information-consuming individuals. What is little better than an accident from the point of view of the individual may well emerge as an expected occurrence from a larger point of view. For while there is only a small likelihood that any accidentally obtained piece of information will be of use to the individual scientist who obtained it, the likelihood that it will be of interest to at least one of his departmental colleagues is much larger. And if enough members of a given department or research group are plugged into branches of the professional grapevine through consultantships, secondary appointments at other institutions, committee services, and personal correspondence and visits, they may collectively be able to assure each of them a good share of the news about work in progress that interests him.

The formal and organized means of communication—especially the periodical literature, including its voluminous abstracting services and review publications—serve the scientist most efficiently when he knows precisely what he is looking for, when he needs the answer to a specific question. When it comes to bringing scientists together with information the significance of which to their own work they have not anticipated; when it comes to pushing out the frontiers, it may be that the system of informal and “accidental” means of communication, inefficient though it may be, is as reliable a mechanism as one can get. In fact, the very frequency of the “chance” occurrences of information transmittal, which was illustrated in the preceding pages, suggests that they may not be altogether accidental; perhaps, if knowledge of a particular item hadn't come to the researcher one way, it would have come through another, although a little later.

One chemist told us of an experience which seems to bear this out rather dramatically. He had done some experimental work in 1955 and had published a report without fully realizing the relevance of his work to the chemical theory of a certain reaction mechanism. Between 1955 and 1957, he was led to earlier literature which suggested this significance of his experiment to him. During the same period, this fact was also brought home to him through three contacts with other scientists which had ensued from his work in three quite independent ways. To what extent one can depend on these apparently fortuitous mechanisms of communication to bring the right combination of scientist and information together is, of course, not known. It is, however, worthwhile to consider the totality of information exchanges among scientists as a system, to accept what appears as “accidental” communication as part of the system, and to examine the ways in which the system, including its unorganized components, may be made to operate more efficiently and more reliably.

FURNISHING ANSWERS TO SPECIFIC QUESTIONS

These pages make it fairly clear that the functions of the scientific communication system extend beyond the bounds of enabling the scientist to get the information which he knows he wants. Yet even to get the scientist the information which he knowingly seeks takes more than the means officially established for this purpose. This can be seen when one examines the ways in which scientists secure the available answers to specific questions. Yet no other function of the scientific communication system has received more solicitous care through formal arrangements than this “reference function.” Copious amounts of planned effort and many specially designed devices—indexes, abstracts, card files, compendia, handbooks, loose-leaf services, and what not—are employed in its service. Analysts study the completeness of coverage of these facilities, the time lags involved in their preparation, the suitability and logical structure of the categories they employ. Scientists and engineers streamline the existing facilities and devise new ones—microfilm libraries, new cataloguing systems, mechanical retrieval systems. These are also the topics which occupy the largest single portion of the program of the present International Conference.

SEARCHES OUTSIDE THE LITERATURE

Here, once again, our exploratory study sought to gain insight by examining the reverse side of the coin. Instances when scientists secured answers to specific questions in other ways than those deliberately designed for this purpose were recorded. As before, it was hoped that this procedure would, on the one hand, illumine the operation of the informal avenues of communication, and that it would, on the other hand, point up the services which the formal reference facilities fail to perform. Eventually, such knowledge may suggest ways of having the formal facilities do more adequately the job they presently fail to do. Or, depending on the circumstances, it might be found more practicable to improve the operation of informal avenues of communication. More of this later.

The following question was included in the interview:

…Can you tell me about the last time you used another channel than just the literature to find the answer to some question that arose in connection with your work?

CONTENT OF INFORMATION SOUGHT

In examining the replies to this question, it is striking how intimately the content of the information sought is tied up with the reason for seeking it outside of the regular channels of the literature search. For in two-thirds of the reported cases, the nature of the information sought either made it improbable that it would appear in the literature at all, or made it seem very difficult to track down, even if published. Most of these searches were for practical details to supplement basic knowledge which was already at hand: unpublished minor details of already published findings; information about the use of techniques and the adaptation of apparatus; quests for the fruits of experience and know-how. For example:

I have two former Ph.D.'s at…Institute. We have some equipment that was developed there, and I called them up for questions about it not long ago.

Or:

Last week there was a conference…in the city…. There were specific questions that were troubling me, about modifications in our instruments. I made it my business to have lunch with some people who were working at other laboratories in the country, and found out if…they had any experience with specific devices that are mentioned in the literature. E.g.,…I asked someone, "Have you used this? Does it work as well as the article reports?"

In the remaining one-third of the episodes, the information was secured through personal channels although the nature of its content would not seem to have barred it from appearing in print, or from being traced if published. But in half of these cases, the information sought had not, in fact, appeared in print at the time it was secured by the interviewed scientists. For example,

I now intend to write someone in Chicago and he will answer me…. A number of things are not available to me [otherwise], and they have it in Chicago…[It is on] the purely theoretical calculations of electronic structure of molecules.

In the other instances the information is known to have been available in print at the time it was secured. Personal channels were used in the following ways: obtaining citations from students; having a friend at a pharmaceutical company arrange for the searching, excerpting, and collating of literature; securing in conversation with a local fellow zoologist the published background information about an organism.

Table 3 summarizes the types of specific information which the interviewed scientists reported having secured through informal channels rather than through a literature search.

TABLE 3. Answers to specific questions sought through personal channels: nature of the information sought

	Number of episodes related by:			
Nature of the information sought	*Total*	*Biochemists*	*Chemists*	*Zoologists*
Publication or indexing unlikely				
Facts to be newly established	2	—	1	1
Practical details on:				
Materials	2	1	—	1
Apparatus	2	—	2	—
Techniques	10	1	2	7
Findings	2	—	2	—
Publication, indexing not unlikely				
Techniques	5	2	1	2
Findings	5	2	2	1
Total	28	6	10	12

PERSONS CONSULTED AS SOURCES

Given that information was sought through personal channels, how did these scientists know to whom to address their questions? How did they decide whom to ask, when more than one possible source existed? Did they address inquiries "cold," or did they tend to seek out colleagues with whom they had already established some sort of relationship? Answers to these questions may shed light on the importance which various forms of access to personal communication hold for scientists. They may also give some insight into what is, perhaps, lacking in the communication picture of scientists in less favored positions, where opportunities of access to experts are not so ample.

Nearly half the inquiries here recorded were addressed to scientists who were easily identified as the ones most qualified to answer the particular question: they were the authors of publications on which clarification was sought, the developers of instruments or techniques regarding which counsel was needed, or the recognized leaders in a specialty.

A few of the inquiries were addressed to individuals uniquely qualified to answer them, although their expertise in the matter was not generally known at the time. How, then, did the inquiring scientists know that it was these particular colleagues who could answer their questions? In each case, it was a more or less fortuitous circumstance which established the contact. One illustrative case is that of a chemist:

I wrote to Professor X, of the physics department at…University, for some information on…about four months ago. It was factual information. I knew the answer was not in the usual literature, but to salve my conscience, I looked in the most recent *Physical Review*. I did know that Professor X had done an experiment which would have given him this information as a by-product…. (How had you known that Professor X had done this experiment?) Because he had done it at… [summer laboratory] two summers ago, when he had lived next door to me.

Perhaps the most extreme example of the fortuity in learning who has vital information is the following odyssey of a zoologist.

We had begun using a new technique for measuring the amount of…in the blood, a technique we had first heard of through the literature. But we had problems with it, so we wrote to the man who had developed it—and even called him on the [long distance] telephone.

Up to this point, information about a procedure was sought from its publicly known author. But this attempt to gain the desired knowledge proved unsuccessful, and the report continues:

But that didn't really help, and we heard that he was having trouble with it, too. It's a notorious technique in that it is very difficult to get consistent results. We might have been able to solve the problem if we had kept working on it. But then I went to a meeting in…[a European city], and saw from the meeting abstracts that a man was there to present a paper who had also used this technique. It turned out that he hadn't done the work [himself], but that he had used a modification which had been devised by a biochemist at NYU. So he gave me this man's name at NYU and when I got back we got in touch with him and solved the problem.

It is to be hoped that most searches for information lead to their goal less circuitously, at least when the goal is so close at hand. It is worth noting that the informal source of information, once located, proved very effective. This episode indicates that much would be gained if *finding* the right personal source of information could be made more efficient. It may well be possible to make better *formal* arrangements for locating such *informal* sources of information, even where it would not be practical or economical to have the actual information carried in the formal media.

A third and final category of the inquiries was addressed to scientists who, although qualified to answer, were not the outstanding experts on the subject. They were, however, previously known to the inquiring scientists, and it was, apparently, this accessibility which determined their choice. Thus questions were, variously, addressed to men "who had worked with me here for a year," "who were known to me from previous conferences," "who had been my fellow-students," "with whom I had worked in the same lab previously," and the like.

The various personal sources of information which were used by the interviewed scientists to secure the answers to specific questions are summarized in Table 4.

TABLE 4. Answers to specific questions sought through personal channels: source used

	Number of episodes related by:			
Source to whom inquiry was addressed	*Total*	*Biochemists*	*Chemists*	*Zoologists*
Publicly known top experts	12	3	4	5
Privately known top experts	3	—	2	1
Accessible, not top experts	14	3	4	7
Total	29	6	10	13[a]

[a] One zoologist counted twice because he used both a publicly and a privately known top expert in pursuit of the same question.

SOME QUESTIONS FOR FUTURE RESEARCH

The reader has already been warned that he must not regard the figures in this table, or elsewhere in this paper, as reliably representing the communication behavior among some defined population of scientists, not even that of the 77 biochemists, chemists, and zoologists who were interviewed at one American university. The purpose of this exploration was to formulate problems and procedures for further investigation; a large part of the schedule of questions was developed and modified as the interviewing progressed, rather than uniformly applied; and even as finally used, some of the questions are not regarded as satisfactory.

But even if there were no question of the reliability and representativeness of what is reported here, it is quite obvious that only a scant beginning of an understanding of some problems and potentialities of scientific communication has been achieved. Even the fuller report deposited with the National Science Foundation does no more than to set the stage for more systematic investigation, as, indeed, was the purpose of this exploration.

What are the next questions calling for investigation? A more extensive and systematic survey will first have to correct the shortcomings inherent in any first pilot study. This means the more systematic collection of data in order to confirm or refute what has been suggested above and in our fuller report. It means the observation of the variations in these forms of communication which may accompany differences in the attributes of scientists, their disciplines, and their positions. It means the examination of the prevalence of various communication processes in diverse settings, including especially those where the opportunities for personal communication with top experts are more limited. Only then will it be possible to draw inferences from the multiple cross-classification and statistical analysis of observations in the traditional manner of survey analysis.

A more detailed illustration follows of some of the questions that must, in our opinion, be answered in order to have a sound basis to guide scientific communications policy. Not all the questions listed below can be answered by interview survey methods; some could, it is hoped, be answered by the judgment of appropriate experts on the basis of materials which an interview survey can provide. The paragraph headings which follow denote the issue of communications policy to which the research questions listed are pertinent.

DETAILED QUESTIONS ON THE FURNISHINGS OF ANSWERS TO SPECIFIC QUESTIONS

The following paragraphs refer to the communication function last discussed: enabling scientists to find the available answers to specific questions.

1. Is any action indicated?

In the preceding pages have been reviewed some instances where scientists had need for information on a specific point and obtained it through informal and personal channels rather than through the formal and established means of a literature search. Various courses of action are conceivable which would either make feasible the securing of information of this type through the formal channels, or else would make the informal channels more effective and more generally accessible for this purpose. Any such action, however, will recommend itself only insofar as the kind of information now secured outside the formal channels is really essential, and is not now adequately accessible to all who seek it.

If any doubts exist as to these points, some empirical tools are available for resolving them. To examine how essential the information in question is, it may be useful to collect, more systematically and extensively than here, a sample of items of information secured outside of the formal machinery of literature searches. This could then be submitted to a panel of experts for judgment as to the importance of the messages. In those cases where the information was to appear in print later, the experts may judge whether the time saved by using personal contacts was essential.

To examine empirically how adequately the information in question is now available to those who seek it, it would be useful to ascertain what is done by scientists who lack access to personal sources of information when they have call for information comparable to that obtained through personal contact by the scientists interviewed here. One would also want to know how often a need for information was felt by a scientist but remained unfulfilled because he knew of no personal way of pursuing it.

2. Should more varieties of information be made securable through the literature, or should informal channels be made more widely usable?

The selection of additional questions for research must be a function of the possible courses of action among which a choice is to be made. The possible courses of action fall into two classes: *(a)* those designed to make the printed media carry more messages of a given kind in such a way that they can be located on demand; *(b)* those designed to make useful personal communication more widely accessible.

If the number of potential users of each of the items of information in question is very large, the printed media will have to be emphasized as much as possible, since personal communication imposes an additional cost in terms of source's time for each additional consumer of a message. If, on the other hand, the number of potential users of any one of these messages is moderate, then one may wish to put more emphasis on enhancing the effectiveness of personal communication. The latter is the only course for those types of information which cannot be economically handled in print. (Cf. Item 6 below.)

To form a judgment of the number of potential users of a message goes beyond the scope of survey research. A survey can, however, give a more accurate picture of the *kind* of messages that are under present conditions diffused through personal channels, and on the basis of this picture it may be possible for experts in the sciences to make some judgment of the potential number of their users.

3. Should informal channels be made more widely usable by enhancing their general accessibility, or by making the likely sources of particular information easier to identify?

Policies which may enhance the effectiveness of personal contacts as channels for securing the answers to specific questions again are of two kinds.

(a) All the many steps which might be taken to encourage the free and frequent give-and-take among scientists, or, perhaps, among specific kinds of scientists, are relevant here. The possible steps are very diverse and include, for example, exhortations to make mail inquiries, the encouragement of inter-institutional visits, the scheduling of teaching duties so as to leave some days open for travel, the arrangement of small conferences on limited topics, and also such long-range policies as the geographic location of research centers in such a way as to enhance opportunities for personal meetings among scientists from different institutions.

(b) An entirely different approach is to enhance the effectiveness of existing personal contacts by enabling scientists to find out speedily to whom best to address a given question. One possible step in this direction might be the regular publication, in a newsletter or in a column in existing periodicals, of very brief announcements of work in progress. (Compare the *Mouse Newsletter* or the *News Bulletin* of the Society for Vertebrate Paleontology.)

As a guide to the emphasis to give to these two approaches, future research should ascertain to what extent the effectiveness of personal communication for reference purposes is now blocked by ignorance of the right source, or by its inaccessibility. This means seeking out the instances where personal communication at present *fails* to perform this function, in order to see to what such failure is due; and if it is due to inaccessibility, what are the main existing blocks.

4. Should more information be made securable through the literature by having more of it printed, or by making that which is printed easier to find?

One can make the printed media more fully perform the reference functions which are now performed by personal communication by *(a)* getting more of the pertinent information into print or *(b)* making such information, once in print, “findable” when it is needed. Research should therefore ascertain to what extent failures to secure information of certain types through the literature are due to its not being published, or to the difficulty of locating it at will even when published.

5. What makes published information hard to locate?

If the material is indeed hard to locate on demand even when it is published, then further research should examine the reasons for this. Earlier pages have shown that, within the limits of this pilot study, the information which is most frequently secured through informal means concerns practical details on the use of techniques and apparatus. Can such items, even when they are published, be located on demand? They would ordinarily appear as incidental entries in the course of research reports which are titled and placed according to their subject matter and not according to procedures used. Do indexing and abstracting services catalog procedural items which appear incidentally in the course of reports on research findings? Do, in fact, suitable and generally familiar categories for cataloguing such information exist?

Survey research can answer only some of these questions. It can also provide materials which may help other experts answer further questions.

6. Why is information of certain types seldom published?

If it is the case that certain classes of essential information are customarily not published at all, the reasons for this should be ascertained. Perhaps these reasons can be inferred from the nature of the information that remains unpublished, once this has been determined. If not, it may be necessary to interrogate the originators of items of information about the reasons for their failure to publish them. Some possible reasons are:

(a) The information is felt difficult to verbalize economically. This might be due to the absence of an adequate, standardized, and generally recognized vocabulary. There might then be a task for semanticists. Survey research could tell the semanticists what type of information is at present felt to be ineffable in this sense. No doubt some information would still remain hard to convey without mutual discussion, or without "being shown." Where that is the case, making the personal channels of communication more widely usable is the only recourse.

(b) Standards of publication make communicating information of certain kinds through the printed word unduly lengthy and laborious. It is possible that standards of publication which are necessary in the reporting of scientific findings are not appropriate to the reporting of ancillary matter. This again could not be judged by survey methods, but may be judged by experts on the basis of material collected through a survey.

(c) Authors or editors believe that few scientists would utilize the information. If a

representative survey should show that the kind of information in question is actually in heavy demand, this fact might persuade authors or editors to change their policy accordingly. More importantly, if steps can be taken actually to increase the utility of such information in published form, for example, by making it easier to locate once published (cf. Item 5 above), the incentives for publishing such information would increase correspondingly.

(d) The information is felt to be "trivial" in some inherent, almost esthetic sense, regardless of its utility. Perhaps it is felt that publishing some kinds of information is lacking in dignity. A possible solution for this problem would be the creation of a special facility for the exclusive publication of information of the kind in question, e.g., information on details of techniques, procedures, and materials. For it is very likely that information that appears trivial when juxtaposed with items of greater theoretical significance in a regular scientific periodical, would lose its felt triviality when published in a special medium. It would then be read by scientists when they are in search of precisely this kind of information; it would no longer constitute an unwelcome interruption of reports on more fundamental topics. A special medium for this purpose could be a special journal, or a special section of existing journals.

DETAILED QUESTIONS ON KEEPING SCIENTISTS UP TO DATE

The following paragraphs refer to the ways in which scientists are kept abreast of current developments in the research areas which are relevant to their own work in one way or another. Some of the issues that seem to face scientific communication policy in this regard and some of the pertinent questions for further investigation will be outlined. Three main problems appear: (1) to assure that news will reach the scientists it should reach; (2) to reduce the labor and time scientists must invest in keeping up; (3) to increase the promptness with which scientists will hear of current developments.

The problem of promptness will not be separately treated here, but many of the problems which it poses for behavior research are identical with questions listed below in other contexts.

A. RESEARCH QUESTIONS CONCERNING THE WIDER DISSEMINATION OF SCIENTIFIC NEWS

In principle, there are three types of strategy for assuring wider dissemination of scientific news: (1) to have more scientific news covered by the literature and other formal information-disseminating facilities; (2) to make the information which is covered in these media reach more scientists; (3) to make the networks of person-to-person dissemination of information about current developments reach more widely. Consequently, the following questions are posed for future research:

1. Does any significant part of information about current scientific developments fail to appear in the literature?

Information which scientists found relevant to their work after obtaining it through personal contacts should be examined to see whether it had been published, or was shortly to be published. If not, steps to make the scientific literature cover additional items of news about current developments may be indicated. What these steps would be cannot be stated without knowledge of the type of information that may now be slighted by the literature and the reasons for its failure to appear in print. The pertinent research questions would be the same as those indicated for an analogous situation in Item 6 above. (No doubt the judgment will be reached that certain types of scientific news cannot be incorporated into the periodical literature, at least not without obstructing its other functions. This is one reason for the continued importance of informal and personal channels of communication.)

2. Why were specific published items of scientific news missed by scientists?

If information about current scientific developments which scientists obtained "too late" or through personal channels was in print at the time, it should be examined further to see why it had not come to the scientists' attention through the literature. Three kinds of reasons may be suggested:

(a) The item fell outside of the scientist's area of attention[5] *as he had defined it.* Where this is frequent, one obvious recommendation is for scientists to define their areas of attention more adequately. This offers little room for action on the part of scientific organizations. They may be able to offer some guidance to scientists if there should be a systematic tendency for the workers in a given field to slight certain types of information.

On the other hand, no scientist will ever be able to define his attention-area so that *all* relevant information will be within its compass. The total area within which news of relevance to him may occur is too vast. Besides, some information becomes "relevant" only in the process of being acquired: a scientist's interests must be molded to a certain extent by what information comes his way. For these reasons, it is not enough to improve scientists' deliberate

[5] In the report from which this paper is excerpted, a scientist's "area of attention" refers to the fields of research with the current developments of which he intends to keep abreast. His "primary fields of attention" are those where he tries to keep up with current developments in detail. His "secondary fields of attention" are those with which he "also needs to keep up to some extent, but not as much."

"gathering" of information; the likelihood that additional information will reach them "accidentally" must be maximized.

(b) The item appeared in publications not regularly scanned by the scientist. Such an item of information will be missed by the scientist even though its topic falls squarely within his pre-defined attention area.

Once again, one possible remedy is for the scientist to adjust his information-gathering practices: he needs to scan additional journals or secondary source material. It is possible that certain portions of the literature are rather generally slighted in this way by one category of scientists or another.

But, as before, no scientist can be expected to scan all the possible primary sources of news that may be relevant for him. If scientists frequently miss news items that appeared in journals which they do not scan, it may be more practical to improve their coverage in secondary media—indices, abstracts, reading lists—than to adjust the scientists' reading habits, especially with regard to journals that only intermittently carry news of relevance to the scientists concerned.

It would be important to know whether any substantial number of scientists misses the *same* items of information due to their appearance in journals which they do not scan because they carry news of interest to them too infrequently. Such a situation might call for new secondary media which would list, excerpt, or reprint selected material from these journals for the benefit of workers in certain specialties.

(c) The item was concealed in the context of an article on another topic. Such an item could be missed even when it does fall within the scientist's pre-defined attention area and appears in a journal which he regularly covers.

Where that is the case, the proper remedial action would seem to lie in the provision of better titles, subtitles, prefatory abstracts, or whatever other cues scientists may use to select papers for reading. It would again be important to know whether there is some general tendency for certain types of information to be frequently missed in this fashion. This would make possible the formulation of concrete recommendations for the more appropriate titling of papers; and if the tendency is of great prevalence, there might even be call for reprinting the relevant portions of these papers in places where they would be more conspicuous.

3. In what fields is published information most likely to be missed in the course of scanning?

Generally speaking, any steps that will lighten the burden and shorten the time required for the scanning of a given portion of the scientific literature will free scientists to scan additional portions. Possible action toward this end is discussed below.

More specific action may be feasible if it is possible to delineate particular areas of knowledge as the likely loci of "missed" information. Scanning is probably least thorough where it is least efficient: with respect to fields of research which only occasionally give rise to information of interest to a given scientist. It is in a scientist's secondary fields of attention that he is most likely to miss some relevant information as he scans the current output. Perhaps publications could be created which would selectively reprint, excerpt, abstract, or index those aspects of a given area of research which would be of interest to specialists in another area. This calls for ascertaining what aspects of the information produced in a given field are of relevance to the workers in other specialties.

4. What are the forms of personal communication which bring relevant scientific news to those who have access to them?

For a variety of reasons inherent in the nature of basic research work, personal communication will, no doubt, have to continue to supply much of the important news to scientists. (In part, it does this by calling published information to their attention.) In order to foster the operation of personal networks of communication with the requisite discrimination, it is necessary to know what sort of personal communication would most profitably link scientists in given specialties and positions.[6]

It is necessary to recognize the most fruitful *occasions* for person-to-person exchanges (conferences, corridor conversations, visits, etc.); the *positions* whose incumbents can become nodal points of information-exchange (consultantships, service on award committees, editorial duties, etc.); and the nature of *relays* through which information may be usefully passed on and shared (through friends at other institutions, through contact with "good correspondents and readers," through departmental colleagues who return from conferences and visits, etc.). Because the usefulness of personal communication differs from discipline to discipline, and possibly from specialty to specialty, the factors which determine its differential utility must be taken into account.

5. What is the present opportunity for scientists in varying positions to have access to the fruitful forms of personal communications?

The scientists among whom this study was carried out probably have easier access to useful personal communication with other scientists than their colleagues

[6] Some possible steps to promote personal interchanges among scientists are discussed in the report submitted to the Foundation.

in most other institutions. It is important to know how much access to such communication scientists in various institutions, professional positions, and geographic locations now have, if plans are to be made to have more of them "hooked up" in useful ways with the network of informal information flow.

B. RESEARCH QUESTIONS CONCERNING REDUCTIONS IN THE TIME AND LABOR OF SURVEYING CURRENT DEVELOPMENTS

Great amounts of time and effort are required of scientists who wish to keep abreast with current developments. Three features are chiefly responsible for this situation: (1) the number of different publications which must be surveyed to achieve satisfactory coverage is very large. (2) the screening of papers to read is time-consuming, if one does not wish to risk missing too many useful items of information. (3) The assimilation of the content of many papers demands considerable additional time and application.

Any reduction in the difficulty of screening papers must come from editorial policy which would provide appropriate clues in the titles or other features of published articles. The most useful clues are probably those which are present in the papers which scientists find easiest to screen. This leads to the following empirical question.

6. What clues are lacking from articles which must be examined closely before their pertinence can be determined?

As scientists scan the contents of a journal, they decide to skip some articles on the basis of easily visible clues, and they examine others more closely. Are there many in this latter group which are eventually found not to have been worth reading? If so, what are the clues which are lacking? This may suggest improvements in the provision of titles, subtitles, prefatory abstracts, or other clues, as a matter of editorial policy.

There are several possible strategies for reducing the number of journals a scientist must scan. One is the greater utilization of secondary source publications for scanning purposes. In order to be useful for this purpose, such publications would have to follow original publication much more promptly than the standard abstracting services are able to do. Special abstracts, lists of titles, or even reproductions of the tables of contents of journals may be considered. Most services of this sort which now exist attempt either to cover the entire output in a given line of work, or to report on its most important segments. It might be more practical to have them report to the workers in a given specialty only that part of the output which appears in journals which they find least worth scanning directly.

A second possible way to reduce the number of different journals a scientist must cover would be the creation of new journals which would publish, reprint, or excerpt articles on topics which cut across the classification of topics according to which journals presently specialize.

A third way would be to institute more sharing of the burden of scanning among colleagues. Basic researchers, at least those interviewed, seem not inclined to entrust to anyone but themselves the actual reading of articles; in this they differ from professionals in certain applied fields, e.g., medical doctors. These researchers may, however, be willing to entrust to each other the scanning of journals, even if not the actual reading of the papers selected. Some of those interviewed do this now.

The choice among these strategies and the further specification of any one of them calls for answers to the following questions:

7. At what level of efficiency is the scanning of each journal or other medium performed?

How many articles or pages must a scientist scan in a journal for each one that he eventually reads? or for each one that he eventually finds to have been worth reading? This, it should be noted, is not simply a question about each journal, but rather about the relationship of each journal to each specialty. For a journal which is read from cover to cover by the scientists in one field may contain only one useful article in ten for those working in another.

For any one group of scientists, it would be useful to classify the journals which they scan according to the proportion of their content that is eventually found worth reading. One may then concentrate on those journals which scientists in a given specialty scan with the least efficiency. The information they secure in this inefficient way may be capable of being made available to them in a more efficient manner. This will require answering still another question:

8. What is the nature of the information which scientists secure in the media scanned with least efficiency?

Perhaps this information can be identified by its content; for example, it might be information about the handling of materials or organisms. Or, it may have to be identified in terms relative to the specialty pursued by the reading scientists. In either event, the nature of the information which is secured in the media scanned with low efficiency may suggest utilizing one of the specialized media which were mentioned in Item 6. The number of scientists who share identical difficulties will have to determine the proper course of action.

C. RESEARCH QUESTIONS REGARDING SCIENTISTS MOTIVATION TO BE INFORMED

Nothing has been said so far of a topic to which behavior research has been fruitfully applied in other communication situations: that of motivating the individual to keep himself informed. This is, for example, a serious problem in the case of medical practitioners, who can devote time to following the news only by resisting strong competing pressures from other professional obligations and from private interests as well. With basic researchers like those with whom we dealt here, no such problem seems to exist: being well informed is recognized as one of their chief responsibilities both by themselves and by their institutions. It remains to be seen whether this is true, for example, for scientists engaged in applied work in industry and elsewhere. It is also possible that a problem of motivating the scientists to keep up with certain types of information exists, even though there is no *general* problem. For example, if it should be felt that some scientists do not take interest in a sufficiently broad span of current developments, the wise assignment to them of certain duties may provide a remedy. Many of the scientists interviewed asserted that their teaching duties caused them to keep up with a much broader range of developments than they otherwise would, and that this had important consequences for their research work. Others related how their interests and attention areas had been broadened by editorial duties, by the writing of books, and so forth. This makes the relation between the breadth of a scientist's attention area and his positions and activities a potentially useful question for research.

The preceding pages have enumerated some questions for further research and have indicated some possible lines of action for improving scientific communication that might recommend themselves, depending on the answers obtained. These examples of possible action are not, of course, to be regarded as recommendations at this point. Additional research questions are listed in the report submitted by the Bureau of Applied Social Research to the National Science Foundation.

APPENDIX. INTERVIEW SCHEDULE

As explained earlier, the interview schedule of questions was continuously modified and developed as the pilot study proceeded. What follows is the text of the most recently duplicated version. Even this is not recommended as an instrument for future work. Aside from certain needed improvements, the schedule was designed for a first exploration. Many of the questions were left open-ended and flexible, and the interviewing was entrusted to members of the Bureau's professional research staff. Two matters require special explanation:

(1) The space provided after each question is not proportionate to the length of the answers that were expected and obtained. Two different methods of recording answers were used. In the case of structured and short-answer questions, lines or boxes for entering the answers are provided on the schedule form. In the case of all other questions, the symbol ### directed the interviewer to record the answer verbatim on separate note paper. These answers, actually constituting the bulk of the interview protocol, were then transcribed onto notched cards.

(2) Interviewers were instructed to ask supplementary questions as needed to round out responses to the questions provided on the schedule. In particular, whenever a question called for the retelling of an actual episode of information transmittal, the word RECONSTRUCT, printed on the schedule form, instructed the interviewer to follow a detailed prepared list of questions in order to obtain a complete account.

NO.__________________

COLUMBIA UNIVERSITY

BUREAU OF APPLIED SOCIAL RESEARCH

1.1 What is your field of specialization? ______________________

1.2 Is there a more specific field with which you are identified?

____

____

1.31 Had you planned to specialize in (1.2)__________ ever since you finished graduate school?
Yes ( )
No ( ): 1.32 Had you planned to specialize in__________(1.1) ever since you finished graduate school?
Yes ( )
No ( ): 1.33 Why and when did you switch? ###
IF "Yes" TO EITHER: 1.34 If you started over, would you choose the same specialty again? ###

1.4 Is this a growing field?
A. Would you say it is given enough attention by [biologists] [biochemists] [chemists] today? ###

1.51 Are you the only one interested in__________(1.2) at [this university]?
Yes, I'm the only one ( ): 1.52 How about in__________(1.1)?
Yes, only one ( ): SKIP TO 1.6
No ( ): Who else?
NO ( ): Who else?

____

____

1.6 Do you expect to take up work in any new specialties in the foreseeable future?
No ( )
Yes ( ) In which?________________________
Would you still continue with your present specialty too?

____

Now I will ask you about your present activities.

2.1 How do you divide your time between teaching, research and administration?

Teaching *Research* *Administration*

2.2 Do you supervise any Ph.D. theses?
No ( ) Yes: How many at the present time?________________

2.3 Do you have any responsibilities away from this campus?
No ( ) (IF NOT MENTIONED)
Yes ( ) Do you act as a consultant any place?

Institution *Position or Duties* *Frequency of attendance*

2.41 Do you have any editorial duties on a scientific journal or series?
No ( )
Yes:
Journal *Position*

____
____
____
____

2.42 Do you referee papers for any (additional) journals? Yes ( ) No ( )

2.43 How much time do you spend on all these editorial duties?________per____

2.5 Do you hold office in a professional society at the present time?
No ( )
Yes: *Office* *Society* *Time spent per month*

____
____
____

2.6 Have you been away on any visiting appointments in the past two years?
No ( )
Yes: *Institution* *Dates* *Position*

____
____
____

2.7 Did you do any work away from [this university] during the past two summers?

No ( )

Yes: Summer 1956 at____________________

Summer 1955 at____________________

We are interested in how a scientist's opportunities to be informed are affected by different parts of his work load. E.g.:

2.8 (AS APPLICABLE:) Do you sometimes hear something relevant to your scientific work through (administrative activities) or your activities (as a consultant) (as an editor) (as________of the______________society)?

No or rarely ( ) Yes ( ) ###

2.9 Does the fact that you teach alter your gathering of information in a way that has bearing on your own research?

A. Has your teaching ever caused you to become acquainted with something that turned out to be relevant to your own work? ###

B. Do students ever call your attention to something you might have missed? ###

} RECONSTRUCT

2.0 If it were up to you would you rather do more or less teaching than you actually do? More ( ) Less ( ) As is ( )

Now let us turn to your research work.

3.1 What research project are you currently engaged in?

____

____

____

(IF NONE:) When did you complete your last project?__________

(IF BEFORE SEPT. '56, SKIP TO 3.5)

(IF SINCE SEPT. '56): What is it about? RECORD ABOVE.

3.2 Are (were) any colleagues working or consulting on this project with you?

No ( )

Yes:

Specialty	*Institution*	*Name (if own University)*	*Responsibility on project*

3.31 Could you tell me about your experiences in securing information that has been needed *for your work* during the last year or so?

Of course, the literature is the standard repository of scientific knowledge. But have you, in addition, used any other channels of information in order to find things you needed to know for your work during the past year?

No ( ): 3.32 Would it have been a handicap if you could have gone only to the literature in order to find things you needed to know for your work?

No ( ) Yes ( )

RECORD COMMENTS AND SKIP TO 3.4 ###

Yes ( ): RECORD COMMENTS

3.33 Can you tell me about the last time you used another channel than just the literature to find the answer to some question that arose in connection with your work?

No or can't recall ( ) Yes ( ): RECONSTRUCT ###

3.4 How about when you weren't actually searching for the answer to some specific question—have there been any instances during the past year where some unlooked-for piece of information came your way that turned out to have bearing on your work?

No or can't recall ( ) Yes ( ): Tell me about the last time this happened. RECONSTRUCT ###

3.5 Sometimes a scientist doesn't hear about some relevant prior work in time. In your own experience, has some knowledge that would have made a difference in your work ever reached you too slowly?

No or can't recall ( ) Yes ( ): Tell me about the last time this happened. How did you finally hear it? ###

3.61 Scientists usually try to keep themselves informed of work in a set number of areas, but from time to time they may start to follow what goes on in some line of research which they had not previously paid much attention to.

Could you tell me the most recent instance of this in your experience, and what brought it about?

Can't recall any ( ) Yes ( ): RECONSTRUCT ###

4.11 I will now ask you some questions about your need to be informed of current developments in different fields.

Here (Chart 1) is a list of some subdivisions of ____________ and related sciences. Would you go down this list and circle the figure "1" opposite any field where you try to keep up with current developments in detail.

Feel free to add any fields that may be omitted, or to divide fields that are too broad.

4.12 This card (Chart 2) lists a number of the possible channels which may keep a scientist aware of what goes on in a given specialty.

Thinking now of these fields where you try to keep up with current developments in detail—like ________ and ________ ("1")—which of these channels is the most important in *calling to your attention* the current developments?

Rank:		Col.ii	Col.iii
( )	a. listening to papers	( )	( )
( )*	b. scanning abstracts of meeting	( )	( )
( )	c. abstracts of periodicals	( )	( )
( )*	d. students or assistants	( )	( )
( )	e. own scanning of journals	( )	( )
( )	f. review articles	( )	( )
( )	review volumes	( )	( )
( )	general science journals	( )	( )
( )*	books	( )	( )
( )*	g. conversations with local colleagues	( )	( )
( )*	h. conversations with others	( )	( )
( )	i. correspondence	( )	( )
( )	pre- or reprints } direct from authors	( )	( )
( )	abstracts	( )	( )
( )	j. references in other reading	( )	( )
( )*	k. presentations in seminars	( )	( )

RANK AT LEAST THE FIRST FOUR.

GET FOLLOWING DETAILS FOR*:

b. Including any not attended? ____________________

d. How? ____________________

f. What kind? ____________________

h. Where? ____________________

g. and h. Is ordinary conversation enough, or do you have to ask specific questions?____________________

k. Which seminars or coll.?____________________

4.13 Would these channels of communication have the same order of importance for your keeping up in the remaining fields which you marked "1"?

Yes ( ) SKIP TO 4.21

No ( ) LIST THOSE WHICH DIFFER BELOW AND RANK CHANNELS ABOVE

____________________RANK COL. ii

____________________RANK COL. iii

4.21 What are some fields on this list (Chart 1) where you also need to keep up to some extent, but not as much—(CIRCLE THEM "2").

4.22 Just how does your need to keep up in, say, ____________________ or ____________________ ("2") differ from your need to keep up with the fields you told me about earlier? ###

A. There are so many ways in which one's needs for information about fields may differ, e.g., in coverage, or in promptness, or in the attention to minor labs; I wonder how *you* see the difference in *your* need to keep up in these several fields.

B. Is there a difference in the amount of attention you need to pay to procedures as opposed to findings?

4.23 Does this description of your need to be informed apply equally to the remaining fields which you marked "2", namely ____________________?

Yes ( ) No ( ) CIRCLE THOSE WHICH DIFFER "3"

4.24 How would the channels on this card (Chart 2) rank in calling to your attention the current developments in fields like ______________ ("2")?

Rank:		Col. iv	v	vi	vii	viii
( )	a. listening to papers	( )	( )	( )	( )	( )
( )*	b. scanning abstracts of meeting	( )	( )	( )	( )	( )
( )	c. abstracts of periodicals	( )	( )	( )	( )	( )
( )*	d. students or assistants	( )	( )	( )	( )	( )
( )	e. own scanning of journals	( )	( )	( )	( )	( )
( )	f. review articles	( )	( )	( )	( )	( )
( )	review volumes	( )	( )	( )	( )	( )
( )	general science journals	( )	( )	( )	( )	( )
( )*	books	( )	( )	( )	( )	( )
( )*	g. conversations with local colleagues	( )	( )	( )	( )	( )
( )*	h. conversations with others	( )	( )	( )	( )	( )
( )	i. correspondence } direct from authors	( )	( )	( )	( )	( )
( )	pre- or reprints } direct from authors	( )	( )	( )	( )	( )
( )	abstracts } direct from authors	( )	( )	( )	( )	( )
( )	j. references in other reading	( )	( )	( )	( )	( )
( )*	k. presentations in seminars	( )	( )	( )	( )	( )

RANK AT LEAST THE FIRST FOUR. GET FOLLOWING DETAILS FOR *:

b. Including any not attended? ______________________________

d. How? ______________________________

f. What kind? ______________________________

h. Where? ______________________________

g. and h. Is ordinary conversation enough, or do you have to ask specific questions? ______________________________

k. Which seminars or coll.? ______________________________

4.25 Would these channels have the same order of importance for your keeping up in the remaining fields which you marked "2"?

Yes ( ) SKIP TO 4.3

No ( ) LIST THOSE WHICH DIFFER BELOW AND RANK CHANNELS ABOVE

______________________________ RANK COL. ______

______________________________ RANK COL. ______

IF NO FIELDS ARE MARKED "3": 4.3 And in the remaining fields, would it be correct to say that you do not need to keep abreast?

Essentially Yes ( ) SKIP TO 4.51

Essentially No ( ) CIRCLE "3"

4.31 How does your need to keep up in these fields ("3") differ from the fields you have been telling me about? ADAPT PROBES FROM 4.22 AND 4.23 ###

4.32 What would be the most important channels (Chart 2) that help keep you abreast of developments in ______________ ("3")?

ADAPT 4.24. LIST FIELDS BELOW; RANK CHANNELS ABOVE

____________________ RANK COL. _____

____________________ RANK COL. _____

____________________ RANK COL. _____

4.4 IF ANY ADDITIONAL CATEGORY HAS EMERGED, CIRCLE "4." ADAPT 4.31 AND 4.32 (RANK ABOVE) ###

4.51 When you need an answer to some specific question in ________ ("2") that arises in your work, how do you go about finding it? ###

4.52 Although you need not keep abreast of developments in the other fields systematically, are there some where you frequently need an answer to some specific question that arises in your work?

No ( ) Yes ( ): 4.52 For example? (CIRCLE "5")

4.53 How do you go about finding such answers? ###

4.61 (You have told me of fields where you follow the findings but not the procedures.) Are you sometimes interested in procedures, methods, or apparatus developed in fields where you have little interest in the findings?

No ( ) Yes ( ): 4.62 For example? (RECORD FIELD OF ORIGIN) ###

4.63 How do you manage to hear of such developments? (IDENTIFY IMPORTANT CHANNELS) ###

4.7 Is there anything different about the way in which you keep abreast of theoretical developments and of experiments and findings?

No ( ) Yes: How? ###

You have told me about your work and your informational needs. I would like now to ask some more systematic questions about various possible sources and occasions of information exchange. Let's start with meetings.

5.11 Please tell me what scientific meetings and conferences you have attended during the last 12 months.

Month-Yr.	*Place*	*Name of conference, society or institution*	5.12 *Check if on program*	5.13 *Check one "got most out of"*

5.12 Were you on the program of any of these? (CHECK ABOVE)

5.13 Which of all these did you get the most out of? (CHECK ABOVE)

5.2 Did you take any other trips out of town for professional purposes during the last 12 months?

No ( ) Yes ( ) Where to? Total No. ________

No. of times	*City*	*Institution visited*	*Purpose of visit*

5.3 How often have you had occasion to visit institutions in this area during the last 12 months? This might include local society meetings, seminars, consultations, or just dropping in to chat?

Altogether ________ *per* ________

Name of institution	*Occasion*	*Frequency*

If other parts of this university, record under 6.41.

5.41 About how often do you have occasion to call scientists at other institutions in the city on the phone? ________ per ________

5.42 And how often do other scientists in the city call you? ________ per ________

5.5 Do you ever exchange long distance calls with scientists elsewhere?

No ( )

Yes ( ): For what sort of purpose? ###

5.61 I would like now to take up some details about the annual meetings of societies, like ____________ (SEE 5.13. IF NOT A SOCIETY MEETING, WRITE IN SOCIETY MEETING RESPONDENT GOT MOST OUT OF.)

What did you get out of attending that meeting—taking into account not only the official program, but the all-around value of attending the meeting? ###

IF PRODUCTIVE, ASK FOR INSTANCES AND RECONSTRUCT. IN EITHER CASE, GO ON TO 5.62.

5.62 Could you check on this page (Chart 3) in what ways attending this meeting has been useful to you?

5.63 During your informal conversations at that meeting, you undoubtedly talked about the work of the people you saw. Were conversations also valuable when the talk was about the work of someone not in the conversation?

No ( )

Yes ( ): RECORD COMMENTS AND EVALUATION. WHAT CONNECTION BETWEEN INTERLOCUTOR AND THIRD PARTY? ###

5.64 Were the most productive conversations at the meeting those with old friends, or those with people you had not known personally before the meeting? ###

5.7 Are there any changes you would recommend in the ways meetings are planned and organized?
No ( )
Yes ( ): What? ###

IF NO AD HOC CONFERENCE ATTENDED, SKIP TO 6.11

5.81 So much then for regular meetings of societies. How about conferences on special topics which you attended during the last 12 months—how did their usefulness to you differ from what you have described? ###

5.82 On the chart (Chart 3), what were the chief ways in which you found special conferences superior or inferior to meetings of societies? (MARK + OR — ON CHART 3, RIGHT MARGIN)

6.11 We have talked about meetings and visits to other institutions. Now I would like to ask about other ways in which you keep in contact with scientists. For example, do you ever use letters to seek information or advice from other scientists?
No ( ) Yes: About how often has this happened during the last 12 months? __________

6.12 Do other scientists ever use letters to seek information or advice from you?
No ( ) Yes ( ) About how often has this happened during the last 12 months? __________

6.22 About how many copies of your reprints, preprints, or abstracts do you send out each time? __________

6.23 How many of them go to a regular mailing list? __________

6.24 Do any researchers regularly send *their* reprints, preprints, or abstracts to you personally?
No ( )
Yes ( ): About how many different researchers do this? __________

6.25 About how many reprints, preprints, or abstracts do you receive in a year from their authors (including those sent on request)? __________

6.26 About what per cent of the people you send your reprints to also send theirs?
__________ IF ALL: and v. va.? __________

6.31 Do you attend any seminars, colloquia, etc., in any part of the university?

Which ones have you attended this year?	*About how often do you attend?*	*Main speakers are: (students-colleagues-outsiders-etc.)*

Any informal discussion clubs, lunch groups, or the like?

____

____

REMINDER: How about [other campus]?

6.32 Aside from such scheduled events, what are the main occasions for "shop talk" with your colleagues in the department?

Occasion *Frequency*

____

____

____

____

6.41 Do you have any occasions to meet colleagues at [other campus]?

Occasion *Frequency*

____

____

____

6.42 Do you have any other occasions to "talk shop" with your colleagues in other departments of the University?

No ( )

Yes: What would be the occasions for such conversations? How often each?

____

____

____

6.51 Have you had any conversations with foreign scientists who visited [this city] in the course of the year?

No ( ) Yes: About how many during the last 12 months?

____

6.52 Have you had any conversations with out-of-town American scientists who visited [this city] in the last 12 months? How many? ____________

6.53 Aside from the seminars, have you had occasion in the last year to have conversations with fellow-scientists from this area who come up here on the campus?

No ( ) Yes: About how often during the last 12 months?

____

6.7 You've now given me a picture of your opportunities for personal interchanges with other scientists. I'd like to ask you some questions about ways in which such contacts may be useful to you as occasions for scientific interchange. Some ways in which conversations might be useful are listed on this chart (Chart 4).

Let me first ask you about conversations you have had in the course of *visits to out-of-town institutions or labs*. When you have such conversations, do they frequently, or occasionally, prove valuable by furnishing answers to specific questions that arise in your work? Please check what applies to your experience on this chart.

6.8 (AS APPLICABLE) How do your experiences in information-exchange at:
_________ (summer station)
_________ (recent previous institution)
_________ (institution of secondary affiliation)
differ from those here? ###

7.1 Who are the 2 or 3 people with whom you most often converse about scientific matters?

____

____

____

7.2 (IF LESS THAN 2 DEPARTMENT MEMBERS NAMED): Who is (next) most frequent among your departmental colleagues?

____

____

7.3 Are there any scientists in your family besides yourself?

Relationship *Discipline* *Check if at* [*this university*]

____

____

7.4 Could you stop to think for a moment of your four best friends—what are their occupations?

Occupation *Check if at* [*this university*]

____

____

____

____

7.5 What discipline, other than your own, do you have the most contact with as a scientist? ______________________________

7.6 Have you done any writing or talking on science to the lay public?
No ( ) Yes ( ): What do you try to achieve by that?
What occasions? ###

7.7 Do you sometimes talk about scientific subjects with non-scientists?
No ( ) Yes ( ): On what occasions?
In what lies the value of such conversations? ###

Finally, we can turn to the reading that you do. (You have already told me something about that but I have some specific questions I would like to ask.)

8.11 Here (Chart 5) is a list of scientific journals. Please check the journals which you regularly scan.

8.12 Are there any other journals, not listed there, which you also regularly scan?
No ( ) Yes ( ): Please write them in on Chart 5.

8.2 So these are the journals which you regularly examine. What is your way of making sure that you see them each time? ###

8.31 Do you regularly scan *Biological Abstracts*? Yes ( ) No ( )
Do you regularly scan *Chemical Abstracts*? Yes ( ) No ( )

8.32 Do you regularly scan: Any other periodic abstracts, bibliographies, indices or reading lists?
any abstracting sections in regular journals?
any annual review volumes?
any book review sections?
No ( ) Yes: What ones?

	CHECK IF			
Name of publication	*Separate Service*	*Abstract in Journal*	*Book review in Journal*	*Volume*

8.33 Do you regularly scan any abstracts of meetings not attended in person?
Yes ( ) No ( )

8.34 Do you use any other facilities or take any other steps in order to know what has come out?
No ( ) Yes ( ): What?

8.41 Would you tell me which are the three most important journals for you to read?

____
____
____

8.42 About what fraction of the articles you actually read appears in these three journals? ________

8.51 Do you find any obstacles in keeping up with advances in your field in foreign countries?
No ( ) Yes ( ) ###

8.52 In what foreign languages can you read the scientific literature?
French ( ) Others: ________________
German ( ) None ( )
Russian ( )
Spanish ( )

8.53 Do you make use of any translating service?
No ( ) Yes ( ): Which? ____________

8.54 Do you find that there is need for more translating services than are now available?
No ( ) Yes ( ): Like what?

____

____

8.6 Do you find review articles or volumes valuable:
(a) in your field?
(b) in other fields? ###
(GET EXAMPLES OF GOOD AND BAD REVIEW PUBLICATIONS AND THEIR FEATURES)

9.11 What would be, say, the five labs or institutions that carry on the most significant work in your field? (RANK IN ORDER OF IMPORTANCE)

____

____

____

____

____

(IF NOT INCLUDED IN ANSWER):

9.12 Where would this university rank?

____

9.13 What "field" did you have in mind when you answered this question (9.11-9.12)?

____

____

9.14 If these five institutions were the only ones with whose work you could be familiar—about what fraction of the work in your field that you actually keep track of would that cover?

____

9.2 As things actually are, how do you manage to keep informed of work at these important institutions? ###

9.3 Generally, is your field the kind where you know pretty well what work most people are doing, or is that not the case? ###

9 4 All told, how satisfactory do you find the opportunities for keeping abreast and for exchanging scientific information?
What are major weaknesses?
Do you have any recommendations for improvement? ###

CHART 1 (BIOCHEMISTRY)

Q. 4.11 In which of these subjects do you try to keep up with current developments in detail?
Identify by circling the figure "1."
(Subdivide or add fields by write-in as necessary.)

Analytical Techniques	1 2 3 4 5
Antimetabolites	1 2 3 4 5
Bio-assay	1 2 3 4 5
Biological Oxidation	1 2 3 4 5
Biophysics	1 2 3 4 5
Biosynthesis	1 2 3 4 5
Blood Chemistry	1 2 3 4 5
Cancer, Biochemistry of	1 2 3 4 5
Carbohydrate Chemistry	1 2 3 4 5
Carbohydrate Metabolism	1 2 3 4 5
Cellular Particles Biochemistry	1 2 3 4 5
Chromatography	1 2 3 4 5
Enzymology	1 2 3 4 5
Genetic Biochemistry	1 2 3 4 5
Hormones	1 2 3 4 5
Immunochemistry	1 2 3 4 5
Ion Exchange	1 2 3 4 5
Isotope Techniques	1 2 3 4 5
Lipid Chemistry	1 2 3 4 5
Lipid Metabolism	1 2 3 4 5
Minerals, Biochemical role of	1 2 3 4 5
Natural Products	1 2 3 4 5
Nucleic Acid Chemistry	1 2 3 4 5
Nucleic Acid Metabolism	1 2 3 4 5
Nutrition	1 2 3 4 5
Photosynthesis	1 2 3 4 5
Protein and Amino Acid Chemistry	1 2 3 4 5
Protein and Amino Acid Metabolism	1 2 3 4 5
Protein Structure	1 2 3 4 5
Radiation, Biological Effects of	1 2 3 4 5
Spectrometry	1 2 3 4 5
Virology	1 2 3 4 5
Vitamins	1 2 3 4 5

Additional subdivisions or fields

______________	1 2 3 4 5
______________	1 2 3 4 5
______________	1 2 3 4 5
______________	1 2 3 4 5
______________	1 2 3 4 5
______________	1 2 3 4 5

CHART 1 (CHEMISTRY)

Q. 4.11 In which of these subjects do you try to keep up with current developments in detail?
Identify by circling the figure "1."
(Subdivide or add fields by write-in as necessary.)

Acid-base catalysis	1 2 3 4 5
Aerosols	1 2 3 4 5
Age determination by isotopes	1 2 3 4 5
Alkaloids	1 2 3 4 5
Colloids	1 2 3 4 5
Cosmo- and Geo-chemistry	1 2 3 4 5
Electrochemistry	1 2 3 4 5
Electrolytes	1 2 3 4 5
Electron Diffraction	1 2 3 4 5
Enzymes	1 2 3 4 5
Explosives	1 2 3 4 5
Fission products	1 2 3 4 5
Free radical reactions	1 2 3 4 5
Heterogeneous Catalysis	1 2 3 4 5
Hydrocarbons	1 2 3 4 5
Infra-red spectra	1 2 3 4 5
Irreversible thermodynamics	1 2 3 4 5
Isotope separation	1 2 3 4 5
Low temperature thermodynamics	1 2 3 4 5
Microanalysis	1 2 3 4 5
Microwave spectra	1 2 3 4 5
Natural products	1 2 3 4 5
Nuclear chemistry	1 2 3 4 5
Nuclear magnetic resonance	1 2 3 4 5
Nuclear reaction	1 2 3 4 5
Oxidation-Reduction	1 2 3 4 5
Paramagnetic resonance	1 2 3 4 5
Photochemistry	1 2 3 4 5
Polymer Chemistry	1 2 3 4 5
Polymerization	1 2 3 4 5
Proteins	1 2 3 4 5
Quantum Mechanics	1 2 3 4 5
Radiation chemistry	1 2 3 4 5
Solid state chemistry and physics	1 2 3 4 5
Steroids	1 2 3 4 5
Theory of solutions	1 2 3 4 5
Thermodynamics	1 2 3 4 5
Tracers	1 2 3 4 5
Ultra-violet spectra	1 2 3 4 5
Visible spectra	1 2 3 4 5
X-ray diffraction	1 2 3 4 5

Mark the following broader categories only insofar as you have not already covered them by *your* markings above. Subdivide them by write-in if necessary.

Analytic chemistry	1 2 3 4 5
Biochemistry	1 2 3 4 5
Inorganic chemistry	1 2 3 4 5
Molecular structure	1 2 3 4 5
Organic synthesis	1 2 3 4 5
Radiochemistry	1 2 3 4 5
Reaction rates and mechanisms	1 2 3 4 5
Statistical mechanics	1 2 3 4 5
Special topics in organic chemistry (Specify:________)	1 2 3 4 5

Any additional subdivisions or fields:

__________________	1 2 3 4 5
__________________	1 2 3 4 5
__________________	1 2 3 4 5
__________________	1 2 3 4 5
__________________	1 2 3 4 5

CHART 1 (ZOOLOGY)

Q. 4.11 In which of these subjects do you try to keep up with current developments in detail?
Identify by circling the figure "1."
(Subdivide or add fields by write-in as necessary.)

Bacteriology	1 2 3 4 5	Paleontology	1 2 3 4 5
Bioassay	1 2 3 4 5	Parasitology	1 2 3 4 5
Biometrics	1 2 3 4 5	Photosynthesis	1 2 3 4 5
Biophysics	1 2 3 4 5	Physical Anthropology	1 2 3 4 5
Cytochemistry	1 2 3 4 5	Protozoology	1 2 3 4 5
Cell physiology	1 2 3 4 5	Sensory Physiology	1 2 3 4 5
Ecology of ________	1 2 3 4 5	Spectrometry	1 2 3 4 5
Electron Microscopy	1 2 3 4 5	Systematic Zoology of	
Embryology	1 2 3 4 5	________	1 2 3 4 5
Endocrinology	1 2 3 4 5	Virology	1 2 3 4 5
Evolutionary theory	1 2 3 4 5	Additional subdivisions or fields:	
Genetics: mechanisms of		________	1 2 3 4 5
mutation	1 2 3 4 5	________	1 2 3 4 5
Genetics: microbial	1 2 3 4 5	________	1 2 3 4 5
Genetics: physiological	1 2 3 4 5	________	1 2 3 4 5
Genetics: population	1 2 3 4 5	________	1 2 3 4 5
Isotope techniques	1 2 3 4 5	________	1 2 3 4 5
Laboratory apparatus	1 2 3 4 5	________	1 2 3 4 5
Metabolism	1 2 3 4 5	________	1 2 3 4 5
Morphology	1 2 3 4 5	________	1 2 3 4 5
Muscle physiology	1 2 3 4 5	________	1 2 3 4 5
Neoplasms	1 2 3 4 5	________	1 2 3 4 5
Neurophysiology	1 2 3 4 5	________	1 2 3 4 5

CHART 2

Q. 4.12 Which of these are most important in *calling to your attention* the current developments in certain fields?

a. listening to papers at meetings and conferences
b. scanning abstracts of meetings
c. periodical abstracts
d. reports from your students or assistants
e. own scanning of journal content
f. review articles and volumes, general science journals, books
g. conversations with colleagues here
h. conversations with scientists elsewhere
i. correspondence, pre- or reprints, abstracts directly from authors
j. references in reading on other subjects
k. presentations in seminars, etc.

CHART 3

Q. 5.62 Ways in which attending meetings of professional societies has been useful to me.

By providing me with:	*Frequently*	*Occasion-ally*	*Seldom or never*
a. answers to specific questions that arise in my work	( )	( )	( )
b. a knowledge of what work is being carried on, where, or by whom	( )	( )	( )
c. scientific knowledge I might otherwise not have learned early enough	( )	( )	( )
d. an opportunity to find out details of other people's work	( )	( )	( )
e. an opportunity to inform others of my work	( )	( )	( )
f. an opportunity to secure useful criticism of my work	( )	( )	( )
g. an opportunity for discussions of broader scientific topics	( )	( )	( )
h. new contacts with scientists which can later be followed up through correspondence or in other ways	( )	( )	( )
i. any other ways?			
______________________	( )	( )	( )
______________________	( )	( )	( )

CHART 4

Q. 6.7 When you have conversations on each of the following occasions, do they frequently, or occasionally, prove useful in any of the ways specified?

	In the course of visits to out-of-town institutions or labs.			*In the course of visits to other labs. and institutions in this area*			*When you have conversations with colleagues here at the university*		
	Freqtly.	*Occ'lly*	*Rarely or never*	*Freqtly.*	*Occ'lly*	*Rarely or never*	*Freqtly.*	*Occ'lly*	*Rarely or never*
(1) furnishing answers to specific questions that arise in your work?	( )	( )	( )	( )	( )	( )	( )	( )	( )
(2) acquainting you with work carried on elsewhere?	( )	( )	( )	( )	( )	( )	( )	( )	( )
(3) enabling you to communicate your work to others, and to secure useful criticism?	( )	( )	( )	( )	( )	( )	( )	( )	( )
(4) providing an opportunity for the discussion of broader topics?	( )	( )	( )	( )	( )	( )	( )	( )	( )

	When out-of-town scientists visit here			*When scientists from this area visit here*		
	Freqtly.	*Occ'lly*	*Rarely or never*	*Freqtly.*	*Occ'lly*	*Rarely or never*
(1) furnishing answers to specific questions that arise in your work?	( )	( )	( )	( )	( )	( )
(2) acquainting you with work carried on elsewhere?	( )	( )	( )	( )	( )	( )
(3) enabling you to communicate your work to others, and to secure useful criticism?	( )	( )	( )	( )	( )	( )
(4) providing an opportunity for the discussion of broader topics?	( )	( )	( )	( )	( )	( )

CHART 5 (BIOCHEMISTRY)

Q. 8.11 Please check the journals which you *scan* regularly.

( ) Acad. des Sciences, Paris, Comptes rendus
( ) Acta Chemica Scandinavica
( ) Am. Chem. Soc., Journal
( ) Am. J. of Human Genetics
( ) Am. J. of Physiology
( ) Am. Scientist
( ) Analytical Chemistry
( ) Anatomical Record
( ) Arch. of Biochem. & Biophys.
( ) Atomic Scientists, Bull.
( ) Bacteriological Reviews
( ) Biochemical Journal
( ) Biochimica et Biophysica Acta
( ) Biological Bulletin
( ) Cancer Research
( ) Chemical Reviews
( ) Endocrinology
( ) Exptl. Cell Research
( ) Faraday Society, Trans.
( ) Federation of Am. Soc. for Exptl. Biol., Proc.
( ) Growth
( ) Hoppe-Seyler's Z. für physiologische Chemie
( ) Inst. Pasteur, Annales
( ) J. of Bacteriology
( ) J. of Biochemistry (Japan)
( ) J. of Biological Chemistry
( ) J. of Biophys. & Biochem. Cytology
( ) J. of Cell & Comp. Physiology
( ) J. of Cell Research
( ) J. of Chemical Physics
( ) J. of Exptl. Biology
( ) J. of Exptl. Zoology
( ) J. of General Microbiology
( ) J. of General Physiology
( ) J. of Genetics
( ) J. of Organic Chemistry
( ) J. of Physical Chemistry
( ) J. of Physiology
( ) National Acad. of Sci., Proc.
( ) Nature
( ) N. Y. Acad. of Sci., Trans.
( ) Nucleonics
( ) Physiological Reviews
( ) Quarterly Reviews (Chem. Soc., London)
( ) Reviews of Modern Physics
( ) Royal Society of London, Proc.
( ) Science
( ) Scientific American
( ) Scientific Monthly
( ) Soc. for Exptl. Biol. & Med., Proc.
( ) Virology
( ) Z. für physikalische Chemie

CHART 5 (CHEMISTRY)

Q. 8.11 Please check the journals which you *scan* regularly.

( ) Acta Chemica Scandinavica
( ) Acta Crystallographica
( ) Am. Chemical Society, Journal
( ) Am. Electrochemical Soc.
( ) Am. Inst. of Chemical Engineers, Trans.
( ) American Scientist
( ) Analyst
( ) Analytica Chimica Acta
( ) Analytical Chemistry
( ) Annalen der Chemie
( ) Annales de Chimie
( ) Atomic Scientists, Bull.
( ) Biochimica et Biophysica Acta
( ) Chemical Engineering
( ) Chemical and Engineering News
( ) Chemical Engineering Progress
( ) Chemical Reviews
( ) Chemical Society (London), Trans.
( ) Chemical Week
( ) Chemistry and Industry
( ) Chimica e l'industria
( ) Deutschen chemische Gesellschft, Berichte
( ) Electrochemical Soc., Journal
( ) Faraday Society
( ) Federation of Am. Soc. for Exptl. Biol., Proc.
( ) Fortschritte der chemischen Forschung
( ) Franklin Institute
( ) Gazzetta chimica Italiana
( ) Helvetica Chimica Acta
( ) Hoppe-Seyler's Z. für physiologische Chemie
( ) Industrial and Engineering Chemistry
( ) Journal für praktische Chemie
( ) Journal of Applied Physics
( ) Journal of Biological Chemistry
( ) Journal of Chemical Education
( ) Journal of Chemical Physics
( ) Journal of Colloid Science
( ) Journal of General Chemistry (USSR)
( ) Journal of Organic Chemistry
( ) Journal de pharmacie et de chimie
( ) J. of Physical Chemistry
( ) Journal of Polymer Science
( ) Justus Liebig's Annalen der Chemie
( ) Monatshefte für Chemie
( ) Nature
( ) Nucleonics
( ) Philosophical Magazine
( ) Physical Review
( ) Quarterly Reviews (Chem. Soc., London)
( ) Recueil des travaux chimiques
( ) Review of Scientific Instruments
( ) Reviews of Modern Physics
( ) Royal Society of London, Proc.
( ) Science
( ) Scientific American
( ) Soc. Exptl. Biology & Medicine, Proc.
( ) Z. für analytische Chemie
( ) Z. für physiologische Chemie

CHART 5 (ZOOLOGY)

Q. 8.11 Here is a list of scientific journals. Are there any among them that you scan regularly?

( ) Acad. des Sciences, Paris, Comptes rendus
( ) AIBS Bulletin
( ) Amer. J. Botany
( ) Amer. J. Human Genetics
( ) Amer. J. Physical Anthropology
( ) Amer. J. Physiology
( ) Am. Mus. Nat. History, Bulletin
( ) Am. Mus. Nat. History, Novitates
( ) American Naturalist
( ) American Scientist
( ) Anatomical Record
( ) Arch. of Biochemistry and Biophysics
( ) Biochemical Journal
( ) Biochimica et Biophysica Acta
( ) Biological Bulletin (Woods Hole)
( ) Cancer Research
( ) Chromosoma
( ) Ecology
( ) Endocrinology
( ) Evolution
( ) Exptl. Cell Research
( ) Federation of Am. Soc. for Exptl. Biol., Proc.
( ) Genetics
( ) Growth
( ) Heredity
( ) Inst. Pasteur, Annales
( ) J. Bacteriology
( ) J. Biological Chemistry
( ) J. Biophys. and Biochem. Cytology
( ) J. Cell and Comp. Physiology
( ) J. Cell Research
( ) J. Exptl. Biology
( ) J. Exptl. Zoology
( ) J. General Microbiology
( ) J. General Physiology
( ) J. Genetics
( ) J. Morphology
( ) J. Physiology
( ) J. Protozoology
( ) Lab. Invest.
( ) National Academy of Sciences, Proc.
( ) Nature (London)
( ) Naturwissenschaften
( ) N.Y. Academy of Sci., Trans.
( ) Plant Physiology
( ) Quarterly Review of Biology
( ) Science
( ) Scientific American
( ) Scientific Monthly
( ) Soc. de Biol. (Paris), Comptes rendus
( ) Soc. for Exptl. Biol. and Med., Proc.
( ) Virology

The Use of Technical Literature by Industrial Technologists

CHRISTOPHER SCOTT

Planning the storage of scientific information in such a way as to optimize the efficiency of retrieval requires, above all, a knowledge of the needs of the user of this information. What kind of information does the user want? How much does he already know about the information he is seeking before he begins his search? Even if he knows his best strategy for retrieval, will he be willing to use it, or will he be held back by habits, prejudices or any other of the innumerable irrational motives which are known to influence human behaviour?

Once stated in this form, the problem is seen to involve the treatment as a single unit of the information storage system and its user. Further, the user has his own storage system, the human brain. Though the storage capacity of this organ for technical information is small compared with that of a library, it is extremely well organized for rapid retrieval. The problem therefore becomes that of the optimum allocation of the load between two very different types of system, one mechanical, the other biological. In solving it, the former system may be regarded as variable, but almost everything about the latter system is virtually fixed, including the nature of its lines of communication with the former system. This may appear pessimistic, but it follows simply from the fact that human brains live in societies with a high degree of inertia. In more concrete terms, any practical plans for changing the mechanical storage system must not presuppose a *radical* change in the habits of the user of scientific information. No doubt over a period of generations he can and will be modified; but if we are interested in the next decade or two we had better take the scientist broadly as we find him and build our system of information storage around him.

It thus becomes of crucial importance to study the working scientist and technologist and to examine the role of scientific information in his work. The present paper is offered as a contribution to this field; it is an account of some

CHRISTOPHER SCOTT Central Office of Information, The Social Survey, London.

results of an enquiry carried out in the United Kingdom by the Government Social Survey for the Department of Scientific and Industrial Research.[1]

DESIGN OF THE SURVEY[2]

One of the primary difficulties in a study of this kind is the wide variety of activities in which different scientists and technologists are professionally engaged. A reasonable homogeneity in the population studied is essential for meaningful results. The present enquiry was severely limited in its coverage, and was thereby enabled to probe more deeply into motives and attitudes. Thus, the population studied was that of *technologists in the British electrical and electronics industry*. This industry was chosen as one employing a high proportion of technical personnel and known to have a high consumption of technical literature. Industrial technologists were chosen rather than theoretical scientists because they are much more numerous and, even if their per capita consumption of technical information is lower, taken together almost certainly account for the greater part of the demand for such information. Data on academic scientists, in some respects comparable with our own on technologists, were obtained by Bernal in a well-known study in 1948.

Despite the above severe limitations of coverage, the population so defined was still very far from homogeneous in its activities, as will be seen below.

The needs of the user of technical information have been examined in many studies in the United States and Britain, but among these a direct approach to the user himself has been comparatively rare. In the present enquiry the user was approached by interview. A set form of questions was employed by trained interviewers, who were individually briefed and tested on their understanding of the instructions. The duration of the interview was up to one hour.

The sample of establishments was limited to those of medium size (200– 1,000 employees). All such establishments in the industry concerned were

[1] Responsibility for the views expressed in this paper is that of the author alone.

[2] The study was designed and supervised by Leslie T.Wilkins, formerly of the Social Survey. The present author is responsible for the analysis and report. Acknowledgment is due in particular to Mr. Saul Herner, who originated some of the questions used in the interview.

The full report of this Survey is the property of the Department of Scientific and Industrial Research. It was too long for submission as a paper to this Conference. In drawing up the present abstract a choice had to be made between giving enough information to allow the reader to judge the validity of the conclusions but limiting the conclusions themselves to a selected one or two, and reporting all the main conclusions with no more than a sketch of the evidence supporting them. As this was an exploratory study and the evidence could not in any case be more than suggestive, it was felt that the second procedure would be of more interest. It must be emphasised, however, that the conclusions cannot at present be regarded as fully confirmed. The most that is claimed is that they are of sufficient interest to warrant the carrying out of further and more intensive studies.

sampled, but the sample was confined to certain regions of the country, which together contained about one-half of the total national population of such establishments.

One hundred and sixty establishments were approached in this first stage of the sampling, of which 127 agreed to cooperate. A response bias in favour of the use of technical literature seems a likely result, though the probable effect on the individual interviewee would be small. A random sample of the technologists within each establishment was selected, giving a total sample of 1,082 persons.

The term “technologist” was broadly interpreted, and covered all persons engaged in research, whether qualified or not, all those with technical qualifications, and all those responsible for planning and development work. A detailed definition of this population was drawn up, and the administration of the establishment selected was asked to prepare a list of all such employees at the establishment. From this list the sample of individuals was drawn by means of random numbers, using a fixed (except for rounding) sampling fraction of 1/40.

Some details of these sampling arrangements were dictated by administrative considerations. It cannot be claimed that the resulting sample strictly represents any “population” in any rigorous proportional way; it does, however, have a wide geographical spread and covers the whole range of technical activities from foreman level to research director.

The principal characteristics of the sample so selected were the following:

1. Median age: 35. Those engaged in management duties averaged about 10 years older than those in research and development.
2. Median time in the present post up to date: 3 to 4 years. Median time in the present firm up to date: 7 to 8 years.
3. Median experience in present category of employment: 8 to 9 years.
4. In the whole sample, 17% held degrees, 22% held some technical qualification but no degree, while 61% had no academic or technical qualifications. Management were about average in this respect. Among research workers 28% had degrees; among the production-supervision group, 6%.

It will be seen that the level of qualifications is low. It can be safely said, however, that every person in the sample was involved in duties to which the possession of technical information would be relevant.

THE FINDINGS

It is hoped to publish the detailed findings of this survey in due course. In the present paper we can give no more than a brief summary.

1. A GENERAL ACTIVITY FACTOR AMONG TECHNOLOGISTS

A large number of the findings of this study can be summarized by listing those characteristics of the individuals in our sample which were found to intercorrelate positively. The list below contains 24 items. Not all the possible correlations between pairs of items were examined; nevertheless a considerable number were, and there are very few items whose inclusion in the list depends on a correlation with only one other item. Thus there is good reason to suppose that if the full 24×24 correlation matrix had been calculated it would contain very few negative values.[3]

Characteristics of technologists found to be positively intercorrelated[4]

BACKGROUND

1. Holds academic or technical qualifications
2. Below average age

READING

3. Sees many journals regularly
4. Has read or scanned many journals in last year
5. Can recall a useful article read
6. Knows of abstracts
7. Would *refer to literature* as first step to get technical information
8. Did *refer to literature* as first step to solve current problem

TENDS TO READ JOURNALS WHICH:

9. Are not strictly within the electrical and engineering field
10. Are difficult
11. Appear at relatively long intervals
12. Contain many pages of text
13. Contain few pages of advertising
14. Contain reports of fundamental work
15. Do not contain review articles, book reviews, etc.
16. Do not contain news of equipment

[3] Item 24 on this list is known to have three small negative correlations with certain of the journal characteristics (10, 12, and 16). Items 2 and 3 are also negatively correlated (because managers tend both to be older and to see more journals). These are the only instances of correlations between two items on the list which are known to be negative.

[4] The brief descriptions in this list do not adequately define the form in which the information was obtained. For fuller information reference should be made to the interview schedule, which is reproduced as an appendix to this paper. Items 1 to 20 in the list are based on the informants' claims. Items 9 to 18 are based on assessments by a qualified librarian of the particular journals which informants claimed to have read or scanned. Item 22 was based on an independent assessment of the "current problem" as described by the respondent.

17. Do not contain social news (appointments, meetings, etc.)
18. Do not contain advertisements for jobs

OTHER INFORMATION-SEEKING ACTIVITIES

19. Attends meetings of technical or scientific societies
20. Attends conferences or courses

GENERAL ACTIVITY

21. Claims to be currently working on a problem
22. Current problem is of a research (as opposed to production or administrative) nature
23. Quotes source external to himself when asked main stimulus for new ideas
24. Recalls useful ideas or information arising from chance event

The existence of a positive relationship between so many variables implies a general factor common to them all,[5] and it seems reasonable to regard this as a factor representing *the degree of activity of a technologist in his work*. If this is so, then it is justifiable to regard a positive position on any of these variables as the *desirable* position, in the sense of the position characteristic of the more active technologist. This gives us a tentative framework within which to judge other results.

2. THE AMOUNT AND NATURE OF TECHNICAL READING

The median number of journals claimed as "seen regularly" was 4 per person. Ten percent claimed to see none. The median number claimed as "read or scanned in the last year" was 6 different journals. The latter is based on a prompt list of 98 of the most widely read journals in the field of electricity, electronics, and general engineering, and we shall use this question as the basis for our further conclusions about readership. The journals listed were classified by a librarian, and journal characteristics were cross-tabulated against reader characteristics. The type of journal found to be read by the more "active" technologist has already been described in the preceding section. It was found that the "better" journals, in this sense, were also the least popular; in other words, the *number* of readers and the *level* of reader bear opposite relationships to the type of journal. The journals thus follow roughly the same pattern as general newspapers and magazines.

The average journal on the list was read by 8% of the sample. The average percentage readership varied, of course, with the journal characteristic. Among those characteristics examined, that which corresponded with the highest readership (15%) was "contains 60 or more pages carrying advertisements

[5] A formal factor analysis was in fact carried out on part of the data and the existence of a general factor demonstrated.

only" (which perhaps shows no more than that advertisers know their business), while the lowest (4%) was "appears less often than monthly."

Respondents were asked to say how thoroughly they read the journals they claimed to see regularly. Did they read at least one article, scan, search the index, or look only at the advertisements? Only a very slight tendency was found for those who saw more journals to read them less thoroughly. This seems to suggest that a technologist's reading is not seriously limited by shortage of time.

Depth of reading was, however, closely related to the source of supply of the journal. Journals paid for by the reader were more likely to be read thoroughly. From these data, and by making some further assumptions, it was possible to arrive at a tentative estimate of the importance to the technologist of the journals supplied to him by his firm. Thus, among journals seen, but not paid for by the reader, in about one-half of the cases he reads at least one article per issue; in about one-sixth he averages less than one article per issue but nevertheless believes that he derives appreciable value from what he does see; and in about one-third he derives little value from the journal and apparently only sees it because he does not have to pay for it.

The place of reading was also asked, and the responses are shown in Table 1.

TABLE 1 "Where do you do most of your technical reading?"

Place of reading	*Number*	*Percent*
At home	643	59
At place of work	288	27
At home or at place of work	67	6
In a library	31	3
In trains	23	2
Others	6	1
Does no technical reading	24	2
Total	1082	100

TABLE 2

Reference	*Number*	*Percent*
Scientific, technical or trade journal	565	52
Advertisements	83	8
Leaflets	37	3
Newspapers	34	3
Books, handbooks	28	3
Abstracts, digests	16	1
Reprints, offprints	8	1
Miscellaneous	2	—
Cannot recall such article	309	29
Total	1082	100

Respondents were further asked: "Can you recall the most recent article in any paper, journal, pamphlet, etc., that was of direct use or special interest to you?" The source of the article was asked, and the answers are shown classified in Table 2. It will be noted that 29 percent could recall no such article.

When a reference was given, the respondent was asked to state how his attention had been drawn to the article. Table 3 shows a classification of the replies. The 309 who could recall no article are excluded.

TABLE 3 "How was your attention drawn to this article?"

	Number	*Percent*
Colleagues within the establishment or firm	149	19
Persons outside	65	8
Persons unspecified	26	3
Reference in journal or book	37	5
Abstracts	28	4
Mass media	5	2
Attention was *not* drawn:		
Searched in literature	139	18
Came across it by chance	324	41
Total	773	100

Note. A response was classified in one of the first three categories only if the *article itself* had been recommended.

These results would appear to have considerable importance. Note first that, despite the wording of the question, more than 40 percent denied that their article had been specifically sought and stated that it had been met with "by chance" in the course of routine reading. This emphasizes the value to a technologist of maintaining a high level of activity in technical reading.

Secondly we see that nearly one recalled article in three was specifically recommended to the respondent by some other person. It seems to follow that personal contacts between technologists should be encouraged to the maximum, and that the value of these contacts lies not merely in the passing on of technical information itself but in the communication of *references to sources* of information.

3. ABSTRACTS, MEETINGS, AND OTHER INFORMATION SOURCES

A series of questions was asked on the use of abstracts. These are shown in Table 4 with the number of persons giving each response.

A considerable majority said that they knew of no relevant abstracts; of those who did know of them many said they did not use them; of those who claimed to use them one-third could give no evidence of having done so recently. We are left with 224 individuals, or 21% of the sample, who were able to give at least one identifiable title of an abstracting periodical in their own field which they had used during the last quarter year. In reality it is probably safe to say that, for the great majority in the sample, abstracts relevant to their work were in existence.

This percentage did not vary significantly between different age groups, but it did depend on the nature of the informant's duties. Thus, for those engaged in research it was 28%, in management 24% and in production supervision 13%.

TABLE 4

Question	*Answer*	*Number*	*Percent of total asked*	*Percent of sample*
Do you know of any abstracts in your special field?	Yes	414	38	38
No	668	62		
Total asked		1,082	100	
If yes				
Do you make use of abstracts?	Yes	335	81	31
	No	79	19	
Total asked		414	100	
If yes				
Which have you used in the last quarter year?	Identifiable title	224	67	21
No identifiable title	63	19		
None	48	14		
Total asked		335	100	

The claim to read abstracts was one of the most diagnostic items in differentiating technologists of high and low activity, in the sense defined in Section 1 above. It is perhaps disturbing, then, to consider that even among those employed in research duties less than half were aware of the existence of abstracts and less than one-third had used them recently.

Finally we inquired how abstracts are used by those who do use them. The 335 persons who claimed to use abstracts were asked: "Do you use these for searches or for news of current developments?" Table 5 shows the responses. We shall be referring to these results in the next section.

TABLE 5 Use made of abstracts

Use	*Number*	*Percent*
Searches wholly or mainly	71	21
News wholly or mainly	142	43
Both about equally	114	34
No answer	8	2
Total	335	100

Attendance at meetings of scientific and technical societies was also examined.

The percentage claiming to attend at least sometimes varied from 62 among those with academic or technical qualifications to 25 among those without, the mean for the whole sample being 40%.

Those who claimed to attend meetings were asked: "Would you say that you obtain a significant amount of information or ideas from these meetings?" Of those asked, a total of 32% said that they did *not* obtain significant information from meetings. This figure rises to 38% among the research group, compared with 24% among the management and production supervision groups. This seems to imply a rather high degree of dissatisfaction with such meetings, particularly among the research workers. It is interesting to contrast this with our previous findings regarding attendance at meetings: that attendance correlates positively with qualifications and many of the other characteristics of the more active technologist. Apparently the active technologist attends meetings, but does so, in many cases, without any faith in their value to him. On our present data we cannot say why this should be. Perhaps he attends from a sense of duty, perhaps to maintain personal contacts, or, more vaguely, to "keep in touch." Or it may be that those who answered negatively would not have denied that meetings were of value to them, but did not regard meetings as a significant *source of information.* It seems obvious that this would be worth further study; whether or not meetings are worth while, they appear to be fulfilling a function at least somewhat different from that which is conventionally supposed.

The survey examined the exposure of the technologist to a wide variety of other sources of information—newspapers, radio, television, exhibitions, books, etc. Here the results are less clear-cut and could not justifiably be quoted without lengthy discussion and reservations.

4. THE ROLE OF THE LITERATURE IN THE TECHNOLOGIST'S WORK

There would be little point in determining how much the technologist reads without also determining whether his reading is relevant to the work he does. To assume that because he reads technical literature during working hours his reading must help him in his work would be naive. It is crucial, in fact, to know whether *(a)* Some technologists read and others do not, irrespective of the technical problem on which they are working, or *(b)* Some problems induce reading and others do not, irrespective of the technologists working on them. These are, of course, two extreme hypotheses, and it seems likely even before we see the evidence that the truth lies somewhere between them; nevertheless they are worth mentioning here as possibilities in order to remind us how little we can take for granted in the present state of our ignorance in this field.

This question was approached in the present enquiry by asking the technologist first to describe the problem on which he was currently working and then to say what was his *first step* in tackling it. Subsequently he described the *second step,* the *third step,* and the *most recent step*. Reference to the literature was given as the first step in 12% of cases, as the second in 5%, as the third in 3%, and as the most recent in 2%. The implications of these figures will be considered later; for the moment we are concerned with examining how the percentage varied with the type of problem and the type of person working on it. For this analysis we limited our attention to the "first step," as the only one in which reference to the literature was common enough to supply adequate data. Since the type of problem and the type of person are, inevitably, related the analysis was complex and cannot be given in detail here. The results were by no means conclusive and it must be emphasized that further confirmation will be required. However, for what they are worth, they suggested that whether reference was made to the literature depended both on the problem and on the problem solver, but rather more on the solver than on the problem.

The problems concerned varied widely—from pure research at one extreme to strictly administrative problems at the other. The fact that this variation had no more effect on reference to the literature than did the variation between people suggests as a possible thesis that the value of the technical literature in the solution of particular problems was by no means self-evident to the technologists themselves.

Returning now to the figures just quoted, we note that reference to the literature as a reported first step in the solution of a problem was far from common. When it did occur, it was nearly always the first step; as a later step it was so rare as to be almost negligible. It is interesting to contrast this finding with the results of another question in which respondents were asked by what

TABLE 6

Source	*Number of responses*	*Percent of responses*	*Percent of persons*
Written material of any kind	625	33	60
Intuition, "thought"—no external source admitted to	436	23	40
Personal contacts (informal)	372	19	34
Observation[a] or experiment	252	13	24
Lectures, meetings—formal contacts	102	5	9
Trade exhibitions	46	2	4
Requirements of job, or of customer	28	1	2
Unclassifiable	48	3	4
Don't know, no answer	24	1	2
Total	1933	100	179

[a] Includes observation of other firms' products, processes, etc.

means they got most of their ideas or stimulation for new ideas on improvements or new methods. Responses are classified in Table 6. (Many respondents gave more than one answer.)

Thus we now find 60% of respondents mentioning the technical literature as one of the most important sources of their ideas. This finding can be reconciled with the results of the former question by supposing that *the main function of the technical literature is not that of a reference source for consultation but a primary source of stimulation*. In other words, the technologist reads for general interest and to keep up to date; only much less often does he use the literature for reference.

This finding, if correct, is of considerable significance, and it would be un-wise to accept it on the evidence of a single pair of questions. Let us examine, therefore, how well it is supported by other lines of evidence.

In Section 2 above we saw that 41% of articles recalled and considered useful had been met with by chance in the course of routine reading. A further 30% were seen as a result of a specific personal recommendation. All those deliberately consulted by our informant on his own initiative in order to find a definite piece of information must be found among the remaining 29%—and it is unlikely that they constitute the whole of this number.

Again, in Section 3 we gave some data on the use and knowledge of abstracts and we showed that only 21% of the sample were able to give the title of an abstracting journal which they claimed to have used during the last quarter year. Since abstracts are the main device for facilitating the consultation of journals, this alone suggests that our informants were not very much concerned with using the technical literature for consultation. But the point is made even more impressively when we look at the data in Section 3 on the purposes for which abstracts were used. Here we found that they were used very much more often for news than for searches. Thus even the principal aid to the consultation of technical literature serves more for news than for consultation.

Finally we may mention some findings on libraries. It was found that more than half of those whose firm had a library did not use it. Of those who did use libraries few were dissatisfied, although by what seem to be reasonable criteria the libraries within British industry are generally regarded as seriously inadequate and very little use is made of external libraries. This suggests that our informants were not really concerned about libraries, and this becomes explicable if they regard the technical literature not as a fund of information to be consulted but as a source of primary stimulation. As long as they see their favourite journals as they appear, either through their own subscription or through the firm's circulation list, they are reasonably satisfied.

There is thus considerable circumstantial evidence for the hypothesis that the literature is used very much more for news than for reference. It is perhaps fair to suggest that anyone familiar with the working methods of technologists and applied scientists will not find this conclusion surprising.

IMPLICATIONS

This study throws some fresh light on the role of the technical literature in the work of the technologist or at least of the technologist working in the electrical industry. It is suggested that the principal role of the literature is to supply useful information *which is not being deliberately sought* by the reader. Compared with this, its role as a reference source is a good deal less significant.

Much work remains to be done to clarify, and put into quantitative terms, the relative significance of these two roles. Nevertheless it is already clear that any approach which takes it for granted that the reader of technical literature is typically engaged in a search for some particular piece of information is seriously out of touch with reality.

At the beginning of this paper it was suggested that the technical reader had certain habits in his approach to technical information, and that any attempt to improve the means of communicating such information to him must be built around these habits as they are, rather than presuppose that they can be changed within any reasonable time span. This claim may have seemed unduly pessimistic, but its intention will now perhaps be clearer: if it is true that the technologist, when reading, is seldom searching for anything, but is reading for whatever he may find, then it seems clear that any improvement in the organisation of the literature for reference will be of relatively marginal value in increasing the amount of communication. Much more might be achieved by contriving that the important material be presented to the technologist in the place where he will see it in his routine reading and in the manner in which it will attract his interest.

If this is so, then the information specialist has perhaps been expecting too much help from the technical reader, for he has seen his job as that of supplying the technical reader with any information he asks for. But, as Bernal has pointed out, his task is much more than that: it is to give the reader the information he *needs,* whether he asks for it or not. Further, if this information is to be conveyed to a reader who is not deliberately seeking it, it is arguable that more attention should be paid to the purely journalistic arts in the presentation of technical literature.

Seen in this light, the two most important tasks are the *choice* and the *presentation* of the material to be communicated to technologists. Many interesting conclusions follow, from which we select one that is particularly obvious: the key position in the organisation of technical information would seem to be that of the editor of the technical journal.

APPENDIX *Interview schedule*

Main Schedule
Firm's Classification Sheet
Record Sheet for Question 13
Prompt List and Recording Sheet for Question 14
Prompt Cards

MAIN SCHEDULE

S.S. 245
Explain purpose of study as instructed

1. Would you describe in about 20 or 30 words (or less) the nature of your work?

Y	X	O
1	2	3
4	5	6
7	8	9

2. Could you classify your work under the following headings?

P	Production supervision	1
R	Management	2
O	Development	3
M	Sales	4
P	Research	5
T	Other (specify)	

. . . .

.	.	6
7	8	9
Y	X	O

3. About how long have you been engaged in____ (as coded above)? ________yrs. ______mos.

4. Can you say by what means you get most of your ideas or stimulation for new ideas, for improvements or for new methods?

Y	X	O
1	2	3
4	5	6
7	8	9

(i) *After reply or*
if no response PROMPT "Is this mainly by"

P	Spoken word (lectures, conversation etc.)	1
R		
O	Written word,	2
M		
P	Just looking round,	3
T	Other ways (specify)?	

4	5	6
7	8	9

(If possible code only one)

5. Are you working on any problem at the moment?

YES	Y
NO	X

If "Yes"
(a) What is the nature of the problem?

		0
1	2	3
4	5	6
7	8	9

(b) How did this problem first come to *your* (personal) notice?

		0
1	2	3
4	5	6
7	8	9

(c) Can you tell me the very first step you took to deal with it?

1. 0 / 1 2 3 / 4 5 6 / 7 8 9

2. The next? 0 / 1 2 3 / 4 5 6 / 7 8 9

3. The next? 0 / 1 2 3 / 4 5 6 / 7 8 9

(d) What steps did you take today/yesterday?

0 / 1 2 3 / 4 5 6 / 7 8 9

(e) Would you classify this as

(i) Y
Use (ii) X
Card (A) (iii) O
(iv) (Specify)

1 2 3 / 4 5 6 / 7 8 9

6. If you want information on a technical problem which you cannot solve from your memory or your usual standard references, what do you usually do *first*?

Y X O / 1 2 3 / 4 5 6 / 7 8 9

7. Here is a list of some items of information you may need in your work. How would you go about getting information on each?

a. An account or description of an apparatus — Y X O / 1 2 3 / 4 5 6 / 7 8 9

b. A standard or specification — Y X O / 1 2 3 / 4 5 6 / 7 8 9

c. A physical or chemical constant — Y X O / 1 2 3 / 4 5 6 / 7 8 9

d. A method or procedure — Y X O / 1 2 3 / 4 5 6 / 7 8 9

e. An established scientific theory — Y X O / 1 2 3 / 4 5 6 / 7 8 9

f. A new scientific theory — Y X O / 1 2 3 / 4 5 6 / 7 8 9

8. Is any person or department specifically responsible for drawing your attention to articles or other publications which they consider you ought to see?

YES Y
NO X

If "Yes"
(a) Who carried out this function?
Position in firm

0 / 1 2 3 / 4 5 6 / 7 8 9

9. Do you feel that you can safely delegate the function of searching journals etc. to a specialist information officer?

Yes, delegate safely Y
No X
Other reply

0 / 1 2 3 / 4 5 6 / 7 8 9

10. Assuming that you cannot cover all the field yourself, do you consider it best to
Use Card (B) (a) Y
(b) X
(c) (Specify)

		0
1	2	3
4	5	6
7	8	9

11. Do you know of any abstracts in your special field?
YES Y
NO X

If "Yes" ask:
(a) Do you make use of abstracts?
Used 0
Not used 1
If "Used" ask:
(a) (i) Which have you used in the last quarter year?

	2	3
4	5	6
7	8	9

(a) (ii) Do you use these for searches or for news of current developments?
Searches wholly or mainly X
News items mainly Y
Both about equally O
(a) (iii) How many actual articles have you referred to via abstracts during the last quarter year?
No........

12. Can you recall the most recent article in any paper, journal, pamphlet etc., that was of direct use or special interest to you?
Reference or description

Y	X	O
1	2	3
4	5	6
7	8	9

Type of item/book, etc.
(a) How was your attention drawn to this article/book, etc.?

Y	X	O
1	2	3
4	5	6
7	8	9

13. Can you now list for me the journals which you see regularly? By "regularly" I mean those which you see nearly every issue? *Go to separate sheet* (*blue*)

14. Here is a list of journals which apply to the electrical industry and the general engineering field. Would you look down and mark off the ones in which you have read or scanned at least one article during the last year? *Go to List* (*green*)

15. Could you say which journals/papers you would have regularly if you did not have more than three?
Three only
1
2
3

16. Where do you do most of your technical reading
P In office/lab./works X
R In trains Y
O In library O
M
P At home 1
T Other place (specify)?

	2	3
4	5	6
7	8	9

17. I should like you to try to remember *All* your reading during the last 24 hours? I will prompt your memory by time intervals.
[Please give title of book; name of newspaper or magazine; or state "general search of the literature on . . . ."]
1. Before leaving home this morning?
R
I 2. On the way to work this morning?
N
G 3. At work this/yesterday morning?

S 4. During lunch hour today/yesterday?
T
A 5. At work this/yesterday afternoon?
R
T 6. On the way home yesterday evening?

7. At home (or elsewhere) last evening?

Use Card (C)	Formal discussions/seminars	9
	Informal/social contacts	10
	Learned society journals	1
	Lectures	2
	Newspapers	3
	Radio programmes	4
	Reference books	5
	Summary publications (Review of recent development in . . . type of pamphlet, etc.)	6
	Television programmes	7
	Text/instruction books	8
	Visits of sales representatives	9

27. That is the end of the formal questions. Perhaps whilst we have been going through these some other ideas have come to your mind. I should like now to make a note of any points you would care to mention.

CLASSIFICATION

Interviewer. . . . .

Authorisation No. . . . .

Day

Mon.	Tues.	Wed.	Thur.	Fri.	Sat.
1	2	3	4	5	6

Date ____________

Code No. of firm ____________

Code of individual ____________

Firm's designation of individual

____

____

Sex		*Status*	
Male	Y	Married	0
Female	X	Single	1
		Widowed	2

Age

Under 25	3
25–34	4
35–44	5
45–54	6
55 and over	7

Has the subject attended a *University or university college?*

YES	Y
NO	X

If "Yes"
What was the major subject studied

____

____

		O
1	2	3
4	5	6
7	8	9

No. of years with firm ____________ yrs.

No. of years in present position ____________ yrs.

Qualifications	*Years when obtained*
____________	____________
____________	____________
____________	____________
____________	____________
____________	____________

Y	X	O
1	2	3
4	5	6
7	8	9

Y	X	O
1	2	3
4	5	6
7	8	9

INTERVIEWER'S NOTES

Use Card (C)	Formal discussions/seminars	9
	Informal/social contacts	10
	Learned society journals	1
	Lectures	2
	Newspapers	3
	Radio programmes	4
	Reference books	5
	Summary publications (Review of recent development in . . . type of pamphlet, etc.)	6
	Television programmes	7
	Text/instruction books	8
	Visits of sales representatives	9

27. That is the end of the formal questions. Perhaps whilst we have been going through these some other ideas have come to your mind. I should like now to make a note of any points you would care to mention.

CLASSIFICATION

Interviewer. . . . .

Authorisation No.. . . . .

Day

Mon.	Tues.	Wed.	Thur.	Fri.	Sat.
1	2	3	4	5	6

Date ________________

Code No. of firm ________________

Code of individual ________________

Firm's designation of individual

____

____

Sex		*Status*	
Male	Y	Married	0
Female	X	Single	1
		Widowed	2

Age

Under 25	3
25–34	4
35–44	5
45–54	6
55 and over	7

Has the subject attended a *University or university college?*

YES	Y
NO	X

If "Yes"
What was the major subject studied

____

____

		O
1	2	3
4	5	6
7	8	9

No. of years with firm ________________ yrs.

No. of years in present position ________________ yrs.

Qualifications	*Years when obtained*
________	________
________	________
________	________
________	________
________	________

Y	X	O
1	2	3
4	5	6
7	8	9

Y	X	O
1	2	3
4	5	6
7	8	9

INTERVIEWER'S NOTES

FIRM'S CLASSIFICATION SHEET 30.12.55

(i) *Name of Firm*
Address *Code No.* . . . . . ☐

(ii) *Staff* Total employed . . . . ☐

Number sampled . . . . ☐

Number of scientists, etc. . . . . ☐

Sampling fraction . . . . ☐

Note. The first entry, all the workers, sets the target for the sample. If the third line (number of scientists and technicians) exceeds sample number, divide number available by number required and enter this fraction against sampling fraction.

(iii) *Main Products of Firm*

Y	X	O
1	2	3
4	5	6
7	8	9

(iv) *Is the Firm*

A contractor (making to own specification etc.)	Y
Sub-contractor (specs. laid down by others)	X
Part of each form of work	O

(v) *Does the Establishment* form part of a firm with more than one establishment?

YES Y
NO X

If "Yes," is it located in or adjacent to establishment forming sampled unit?

YES 0
NO 1

(vi) *Has the Firm a Central Research Unit?*

YES Y
NO X

If "Yes," is it located in or adjacent to establishment forming sampled unit?

YES 0
NO 1

Note. If the central research establishment does not form part of the establishment sampled, only the scientists and technicians in the establishment form the basis of the individual sample—do not include the separate research establishment.

THIS SHEET IS TO BE PINNED FIRMLY TO THE TOP OF THE COLLECTION OF INDIVIDUAL QUESTIONNAIRE FORMS RELATING TO THE FIRM. REFERENCE NUMBERS MUST ALSO APPEAR ON ALL FORMS.

INTERVIEWER'S RECORDING SHEET FOR QUESTION 13

Individual's Code Ref. [ ]

13. Would you please list below the journals, etc., which you see regularly? By "regularly" we mean that you see nearly every issue.

Title of Paper/Journal, etc.	*After listing see below*		
____________	R S I A	R I	m b f C F O P rl
____________	R S I A	N	m b f C F O P rl
____________	R S I A	G	m b f C F O P rl
____________	R S I A	O	m b f C F O P rl
____________	R S I A	N E	m b f C F O P rl
____________	R S I A		m b f C F O P rl
____________	R S I A	O N	m b f C F O P rl
____________	R S I A	L	m b f C F O P rl
____________	R S I A	Y	m b f C F O P rl
____________	R S I A	I	m b f C F O P rl
____________	R S I A	N	m b f C F O P rl
____________	R S I A	E	m b f C F O P rl
____________	R S I A	A C	m b f C F O P rl
____________	R S I A	H	m b f C F O P rl
____________	R S I A	B	m b f C F O P rl
____________	R S I A	L	m b f C F O P rl
____________	R S I A	O C	m b f C F O P rl
____________	R S I A	K	m b f C F O P rl

Note. To assist memory and coding, letters forming code are in italics in the explanation. Only one code is needed in each set.

First block of letters relate to reading. Please circle as applies.

R = *R*ead on average one or more articles in full per issue
S = *S*can actual articles—going over pages
I = *I*ndex, refer to, and look up articles which appear interesting
A = *A*dvertisements are main concern

Second block of letters relates to source of journal.

Your OWN Copies m = *m*embership, received as a right ot
b = *b*ought
f = *f*ree

Loans or References C = *C*irculation list
F = *F*irm's Library—loans not via circulation list (C)
O = *O*ther Library is source of loan
P = *P*rivate individual lends copy
rl = *R*ead in *L*ibrary

PROMPT LIST AND RECORDING SHEET FOR QUESTION 14

Individual's Code Ref: [ ]

14. Here is a list of journals which apply to the electrical industry and general engineering field. Would you look down the list and ring the Y's of the ones in which you have read or scanned at least one article during the last year.

Y X A.T.E. Journal (Automatic Telephone and Electric Co., Ltd.,)
Y X Acoustical Society of America, Journal of the (U.S.A.)
Y X Advances in Physics
Y X American Journal of Physics (U.S.A.)
Y X Applied Scientific Research (B) (Acoustics, Optics, Electrophysics)
Y X Applied Statistics
Y X Atomics
Y X Automation
Y X A.W.A. Technical Review (Amalgamated Wireless Australasia Ltd.)(Austr.)
Y X B.E.A.M.A. Journal (British Electrical and Allied Manufacturers Association)
Y X Bell Laboratories Record (U.S.A.)
Y X Bell System Technical Journal (U.S.A.)
Y X British Communications and Electronics
Y X " Journal of Applied Physics
Y X " Management Review
Y X " Scientific Instruments Research Association, Bulletin of the
Y X " Welding Journal
Y X Communications and Electronics (U.S.A.)
Y X Computors and Automation
Y X Consulting Engineer
Y X Control Engineering
Y X D.S.I.R. Radio Research Special Reports
Y X Discovery
Y X Electrical Communication (U.S.A.)
Y X " Journal
Y X " Manufacturer
Y X " Review
Y X " Times
Y X " World (U.S.A.)
Y X Electrochemical Society, Journal of the (U.S.A.)
Y X Electronic Application Bulletin
Y X Electronic Engineering
Y X Electronics (U.S.A.)
Y X Endeavour
Y X Engineer
Y X Engineering
Y X Factory Manager
Y X Faraday Society, Discussions of the
Y X " " , Transactions of the
Y X General Electric Review (U.S.A.)
Y X Illuminating Engineering Society, Transactions of the
Y X Impact
Y X Institute of Fuel, Journal of the
Y X " " Metals and Metallurgical Abstracts, Journal of the
Y X Institution of Electrical Engineers, Journal of the
Y X " " " " Proceedings of the, Part I (General)
Y X " " " " " " " Part II (Power)
Y X " " " " " " " Part III (Commun.)
Y X " " " " " " " Part IIIA (Commun.)
Y X " " " " " " " Part IV (Monographs)
Y X " " " " Students Quarterly Journal

Y X Institution of Electronics, Proceedings of the
Y X " " Production Engineers, Journal of
Y X " " " " Proceedings of the[a]
Y X Institute of Radio Engineers, Proceedings of the (U.S.A.)
Y X Instrument Engineer
Y X Instrument Practice
Y X Journal of Applied Mechanics
Y X " " Scientific Instruments
Y X Light and Lighting
Y X Machine Design
Y X Machinery
Y X Management Abstracts
Y X " Digest
Y X Materials and Methods (U.S.A.)
Y X Mechanical Handling
Y X National Bureau of Standards, Journal of Research (U.S.A.)
Y X " " " " , Technical Bulletin of (U.S.A.)
Y X National Research Council of Canada Research News (Canada)
Y X Nature
Y X P.E.R.A. Bulletin (Production Engineering Research Association)
Y X Physical Review (U.S.A.)
Y X Physical Society, Proceedings of the
Y X Post Office Electrical Engineers Journal
Y X Power (U.S.A.)
Y X Power Apparatus and Systems (U.S.A.)
Y X Power Engineering (U.S.A.)
Y X Production Efficiency
Y X Radio Research Council, Report of
Y X Research
Y X Review of Modern Physics
Y X " " Scientific Instruments
Y X Royal Electrical and Mechanical Engineers, Journal of the
Y X Royal Society, Philosophical Transactions of the
Y X " " , Proceedings of the (A)
Y X Science Abstracts, Section A (Mathematics, Physics, etc.)
Y X " " " B (Mechanical, Electrical and Civil Engineering)
Y X Scope
Y X Society of Instrument Technology, Transactions of the
Y X Technical Book Review
Y X Television Society, Journal of the
Y X Times Review of the Progress of Science
Y X " " " Industry
Y X Tooling and Production
Y X Wireless and Electrical Trader
Y X Wireless Engineer
Y X Wireless World
Y X Works Management

Do you think there are any outstanding publications in your field which do not appear on this list?

If so, could you name them?

[a] After the prompt list had been drawn up and used it was determined that "I.P.E. Proceedings" did not exist as a separate publication. Responses to this item have been used in all data relating to the *total number* of journals, but have been excluded from data based on *classification* of the journals.

PROMPT CARDS FOR QUESTIONS 5(e), 10, 26

CARD A

Would you classify this problem as:

(i) A long term problem involving only routine methods ... ... Code Y
(ii) A short term problem involving only routine methods ... ... Code X
(iii) A long term problem where insufficient fundamental knowledge exists at the moment ... ... Code O
(iv) In some other way? (Please specify)

CARD B

Assuming that you cannot cover all the reading in your field yourself do you consider it best to:

(a) Have a large number of journals looked through by an information officer and your attention drawn to those he considers you should see? Y
(b) Personally go through as many selected ones as you can? X
or (c) Have you some other solution to the problem of the large amount of printed matter produced? O

CARD C

Could you tell me for each of these if you can remember any occasion when you have obtained useful information from:

Abstracts
Advertisements in trade journals
Articles in trade journals
Catalogues and other sales literature
Conferences
Exhibitions
Films (general exhibition)
Films (special instructional)
Formal discussions/seminars
Informal/social contacts
Learned society journals
Lectures
Newspapers
Radio programmes
Reference books
Summary publications
(Review of recent development in.... type of pamphlet, etc.)
Television programmes
Text/instruction books
Visits of sales representatives

Requirements of Forest Scientists for Literature and Reference Services

STEPHEN H.SPURR

Within the professional field of forestry, there exists a considerable body of scientists employed primarily as research workers with governmental experiment stations and large industrial concerns, and as University teachers. These number approximately 1500 in the United States alone and perhaps as many more around the world.

The duties of these scientists are oriented toward the multiple use of forests for wood production, water management, game management, range management, and human recreation. The shortage of actual data on the growth, development, and behavior of both forest plants and animals is such, however, that much of the actual research can be classified as fundamental. Most "pure" botanists and zoologists have preferred to work with organisms easier to handle and control in the laboratory with the result that their direct contribution to an understanding of forest trees and large forest animals has been somewhat limited.

Our forest scientists must be trained in the essentials of botany, but must also be exposed to various aspects of zoology, geology, soils, meteorology, statistics, photogrammetry, and other disciplines. They are almost invariably less specialized than comparable botanists—their closest relatives in the scientific world—but may well claim to be more liberally educated in that usually they have had a greater exposure to a greater range of subjects and of problems. They are not botanists, but scientists in their own right, and may be termed "forest scientists." Those concerned primarily with the biology of forest trees may be termed "forester-botanists."

The literature of forest science is both voluminous and diverse. The selection of most scientists from men trained as professional foresters at either the Bachelors' or Masters' level has given rise to a scientific body long on competence in the field, but short in competence as library workers. This combination of a

STEPHEN H.SPURR Professor, School of Natural Resources, University of Michigan, Ann Arbor, Michigan.

large literature and a body of scientists poorly equipped to cope with it has created serious problems in forestry research and has tended to limit the scientific productivity of the group.

The present paper is an effort to evaluate the requirements of forest scientists for scientific literature and reference services. The material is largely personal, drawn from the author's experience in preparing two combination text-monographs and more than a hundred other written contributions in the field, and in founding and editing a number of serial publications.

THE PLACE OF LITERATURE IN FOREST RESEARCH

An examination of almost any issue of a forestry journal indicates minimum reliance of most forest scientists upon literature survey.

Taking a single issue of the American professional journal strictly at random (*Journal of Forestry,* April 1951), a check of the ten major articles showed that 3 cited no references, 1 cited only his own earlier work, and the other 6 cited only American writings from essentially the current generation and locality. No historical references, references to other parts of the United States (except to textbooks), and no foreign references were included.

The situation in *Forest Science,* founded to stimulate scholarly work, is better as far as literature citations are concerned. Again taking a single issue at random (September 1957), about half of the 12 major articles indicate that the authors have made a serious attempt to find out what has been done on their problem by others. About an equal number—but not necessarily the same ones—have looked into the Canadian literature if they were Americans (or into American literature if Canadians). The same number cited one or more European references. One even cited the English summary of two Japanese articles—the only citations in the issue referring to other than American, Canadian, or western European work. Even so, the present writer, who has been editor of *Forest Science* since its inception, can state that for most of the articles submitted to this journal, one or more highly pertinent references uncited by the authors can be turned up merely by five minutes perusal of standard abstract and bibliographic works.

We may ask whether this failure to consult the literature with sufficient care and perspicacity affects the quality of current research. It does, and to a very marked extent. There is much to be learned from work on ecologically similar tree species and forest types in other parts of the world, from botanists and others who publish on similar subjects in other literatures than forestry, and from competent men of previous generations who were concerned with similar problems.

Without belaboring the obvious, we may note that forest science in western Europe dates from the middle of the nineteenth century, whereas American and Canadian research became significant only well into the twentieth century. The forest types of western Europe and eastern Asia contain many species closely related to ours and growing on soils and in climates similar to ours, thus presenting many comparable problems. The forest literatures in German, Russian, Swedish, Finnish, and other tongues are substantial and largely pertinent to American research.

Even in the United States, the work of a previous generation of foresters is all too often overlooked, and but little effort is ordinarily made to uncover applicable research published in botanical and zoological journals. Articles published in university and museum serials, in local academy of science journals, and in similar series of limited circulation are as apt as not to be completely ignored, especially if written by relatively unknown scientists.

An indication of what can be accomplished through exhaustive literature survey is given in a later section.

THE NATURE OF FORESTRY LITERATURE

First, however, let us consider and evaluate the present literature relating to forest science.

Relatively little of it is adequately summarized in book form. There are essentially only two complete book literatures in forestry: one American and one German. A number of admirable books on individual subjects have been published in England, France, Switzerland, etc., but one must look to the United States and Germany for texts in most of the aspects of forestry.

Taking American book literature for our analysis, there are currently about two score books in print, largely textbooks for undergraduates published by three New York publishing houses. Of these, certainly less than half are products of men and work that have come into prominence since World War II, and only a portion of these may be said honestly to cover and interpret their fields adequately. In other words, we have in the neighborhood of a dozen or so texts that can be said to be both up to date and comprehensive. These cover perhaps half of the subjects taught in professional forestry schools today.

Monographic treatments in the English language are few and far between. In tree physiology and morphology, we still rely upon a 1929 translation of an earlier German work. In most fields we have not even that. Only one or two American tree species have been adequately monographed in recent years, and only a few segments of our subject matter field.

There is at least one forestry journal published in most countries at quarterly intervals or better, and a good selection of these are available in a dozen or so American libraries. In the United States, we have had a professional journal for more than a half-century, and a scientific journal since 1955. These are generally available and are generally consulted.

A very substantial amount of scientific material in forestry, however, appears not in books or journals, but in minor serials that are distributed at irregular intervals to limited circulation lists. Most of the research of the U.S. Forest Service, for example, is reported in lithoprinted form by one of the ten regional experiment stations and two territorial research centers. Each station usually puts out a number of such serials, ranging from one-page throw-aways to research summaries in annual reports and more lengthy station papers. To maintain a file of these for the United States alone is impractical for most scientists. Even the library files are generally incomplete as few librarians catalog or otherwise keep careful check of this “ephemeral” literature.

In addition to the lithoprinted output of federal agencies, many additional data are included in similar lithoprinted and mimeographed leaflets distributed by forestry schools, industries, and state agencies.

While it must be acknowledged that the best research eventually is published in bulletin or journal form, the fact remains that a substantial portion of forestry literature is more or less buried in many hundred serials, reproduced by duplication processes other than conventional printing, published at irregular intervals, and circulated to small and often unedited mailing lists.

Finally, a substantial quantity of fundamental research is available only in unpublished theses. Fortunately, these are becoming more accessible through the use of microfilms.

CASES IN LITERATURE SURVEY

With this background, we may now investigate a number of specific cases as a means of illustrating the requirements of forest scientists for literature and reference services. These are all taken from the writer's own experience, and represent instances where an effort has been made to evaluate the problem of what can be gotten from scientific literature on a specific topic, and how it can best be obtained.

FOREST AERIAL PHOTOGRAMMETRY

At the end of World War II, the author was active in research involving the application of aerial photographs to the inventory of forests and other natural resources. The program was quite successful in so far as widespread adoption of our innovations and principles was concerned. Most of us concerned with the topic felt that we were in some sense pioneers opening new horizons of applied science. We were aware, however, of the Canadian use of aerial photographs and aerial sketching for reconnaissance purposes since World War I, and we knew that a few papers had appeared on the subject in various other countries.

In preparing a book on the subject, it was astounding to find several hundred references on forest photogrammetry. Among them was a series of doctoral dissertations written in Germany under the aegis of Dr. Hugershoff at the forestry school at Tharandt in Saxony, which, had the dissertations been available in English, would have greatly shortened discussions, arguments, and field investigations in this country in recent years.

Later, a complete bibliography on the subject was attempted (published by the University of Michigan) in view of the fact that existing bibliographies were noticeably incomplete. Almost no help was obtained from the library catalogs of the U.S. Department of Agriculture, Yale, Harvard, and other major American forestry libraries, and only a few additional references were picked up from major western European libraries (Oxford being the best in this regard). Getting to see many of the items themselves involved major activity in interlibrary loans, microfilm purchases, and foreign travel—activities too time-consuming to be carried out under most circumstances.

It may be amusing to note that the antiquity of our supposedly revolutionary postwar research is indicated by a news item uncovered in a Berlin newspaper of 1887 which relates the experiences of a practicing German forester who attempted to map his forest by taking photographs from a hot-air balloon, and anticipated many of the problems that confronted us nearly sixty years later.

FOREST INVENTORY

On a subsequent research project at the Harvard Forest, we were confronted with a problem of the statistical correlations between the volume and growth of forest trees on one hand, and the conventional measurements of tree diameter, taper, height, and ring width on the other. Forewarned by the experience with photogrammetry, we undertook a prior analysis of previous work on the subject.

Choosing spruce and fir in the northeastern United States and adjacent Canada for the test, it was found that more than a score of investigators had assayed volume estimates independently, more or less ignoring the efforts of others attempting the same thing. Not only was the duplication excessive, but the labor involved was also so great as effectively to prevent progress in this field for many years.

In preparing a monograph on this subject, during the course of which a reasonably complete survey of the literature was made, it became apparent that the lack of fundamental research projects based upon a thorough understanding of previous findings had resulted in holding forest inventory techniques to an empirical basis over a period of several decades.

SPECIFIC GRAVITY AND TREE GROWTH

A major controversy in forestry circles has long concerned the effect of growth rate on the specific gravity and other properties of the wood cells. In effect, one group, primarily Americans, have claimed a high correlation, while another group spearheaded by South Africans has countered that there was little or none. In 1950, the present writer initiated an experiment to try and resolve the problem for at least one species of pine in one locality.

At a later date, a routine survey of previous literature was begun. The large number of existing papers turned up prompted an effort to follow the problem back to its beginnings. It was found that Theophrastus clearly implied that the growth rate of trees was used as a guide to the strength properties of wood by the ancient Greeks. A substantial German literature on the subject contributed substantially to an understanding of the relationship. Finally, a detailed study in France around the middle of the nineteenth century was unearthed. The experiments carried on at that time and the conclusions reached had been ignored almost completely by later workers. Nevertheless, when the extensive literature was studied and the experimental data were finally tabulated, it became evident that the conclusions reached a hundred years before in France were still essentially sound. In fact, the whole problem could be resolved quite well by drawing inferences from the data in the existing literature—data that simply had not been carefully studied and digested by many of the later protagonists.

THE NATURE OF THE FOREST COMMUNITY

Although complete surveys of existing literature are possible when a scientist has the inclination and the time, and when the number of previous works are in the tens or the hundreds as in the specific gravity problem, many cases exist when such a survey is impossible. An illustration is provided by the current project of the writer, an effort to integrate conflicting theories and philosophical approaches on the nature of forest communities.

A project of this kind involves at least passing acquaintance with much of the entire literature of forest ecology, as well as with that of certain related aspects of tree physiology, genetics, plant geography, geobotany, soils, and meteorology.

Since 1945 alone, somewhere in the neighborhood of 20,000 items have appeared in scientific literature around the world dealing with forest ecology or closely related aspects of other fields. All these items should be scanned. From them, perhaps 5000 references are directly applicable to the problem. By winnowing these and being ruthless in rejection, it is hoped to limit the final work to from 1000 to 2000 citations.

It is thus manifestly impracticable to read carefully all the literature relating to this subject that has appeared since 1945. Fully as much again was published prior to that date, including many of the basic works upon which the classical concepts and diverse schools of thought are based.

How to proceed? Reading all the originals is impossible, even were it not for the language barrier in many instances. Bibliographies and catalogs are invariably incomplete and out of date. Even were they not, they do not provide the necessary information. The only solution is some form of abstract or record-retrieval system. The survey (and the resulting book) must be developed essentially from abstracts, supplemented by a good deal of time-consuming library work on the most important items. In choosing these key writings, the scientist must rely upon his imperfect memory and notes of his past reading, and upon exposure to the work of others.

Clearly, studies of this type will continue to be limited in numbers and in quality as long as the scientist must rely upon original sources as the primary basis for his information.

EXISTING REFERENCE FACILITIES

Fortunately, considerable progress has been made in the field of forestry for the systematic compilation of scientific information. Without this progress, the study last referred to would be impracticable even today.

A reasonably adequate and detailed system for the classification of forestry literature has been evolved at Oxford, revised by an international working committee, and published in key languages. This system, a decimal expansion of a single library class, covers many pages, provides for cross-indexing, and is accompanied by alphabetical and other indices.

At least three organizations are making a major effort to furnish bibliographic services of current forestry literature. These include the Commonwealth Forestry Bureau at Oxford, the U.S. Department of Agriculture, and the Federal Forest Experiment Station at Reinbek, Germany.

Of these, the output of the Oxford group is by all odds the most important to the English-speaking scientist. Here are published a quarterly *Forestry Abstracts* journal, giving more or less detailed abstracts of several thousand scientific papers each year; and a monthly title card service, providing 3-by-5-

inch cards carrying the bibliographic citation and commonly a very brief abstract. The former service provides abstracts of the more important works, while the latter covers in addition a substantial number of minor or quasi-scientific items, but provides but little detail as to the content of the listed items. Both services are keyed to the Oxford decimal classification as to main class and for cross-indexing.

In the United States, *Biological Abstracts* covers some forestry literature; but its coverage is incomplete, but little subdivided, and poorly cross-referenced. This journal, in the opinion of the present writer, is of relatively slight value or interest as far as its coverage of forestry literature is concerned. As an entree to botanical and zoological literature, however, it is of great value and importance.

Most American forestry libraries do not catalog separately or cross-reference adequately most of the journal articles and minor publications which form the bulk of the world forestry literature. Their card catalogs are thus of relatively little value to the scientist wishing to go beyond the obvious sources. (The forestry library at the University of California is an exception to this.)

There are, however, a number of excellent bibliographies available to forest scientists. The Bradley Bibliography compiled by Dr. Rehder for the Arnold Arboretum provides an almost unbelievable coverage of the literature dealing with woody plants prior to 1900. Furthermore, most of the items may be found in the various libraries at Harvard and Yale (although this in itself requires considerable sleuthing ability). The twentieth century literatures of the United States, Canada, and Germany are adequately summarized in various bibliographies up to relatively recent years. Postwar citations up to the current year may be obtained from the three bibliographic services mentioned above.

It is thus possible to construct a reasonably complete bibliography in a relatively brief time. For instance, American writings on almost any forestry subject may be compiled from the Bradley Bibliography (prior to 1900), the Munns Bibliography (to 1929), U.S. Department of Agriculture and Forest Service sources, and *Forestry Abstracts* (reasonably complete since 1945 or thereabouts). The problem of where to find the references and how to abstract critical information from them remains.

AN APPROACH TOWARD A SOLUTION

The present writer has given a good deal of attention to the development of a literature record that will meet the requirements of the forest scientist. The personal system, evolved at the School of Natural Resources of the University of Michigan since 1952, may serve as an indication of a possible solution in a scientific field where the research workers are still relatively few and the literature is still far from voluminous.

Essentially, the system is a large-card catalog of printed and other abstracts catalogued and indexed in accordance with the Oxford decimal classification. Two copies of *Forestry Abstracts* are periodically disassembled, the abstracts cut out, and filed in accordance with the classification code printed on them. The major items are pasted on index cards, while the minor contributions are filed by separate decimal classes and geographical locality in envelopes of the same size in the same catalog.

The system was evolved to be kept up to date by the scientist himself and his assistant with a minimum of time spent. An average of 1 to 2 man-days per quarter is all that is required. Reprints are kept in a parallel file (standard filing cabinets), following the identical classification scheme. In the course of doing a large part of the filing himself, and in perusing and filing reprints in accordance with the same system, the author has been able to keep reasonably well informed of current publications in his particular field.

Were the system to be modified to an institutional basis rather than a personal one (and this has already been done at another forestry school), obvious additions and modifications could be made to evolve a sound scientific information catalog. The following points would be considered by the writer were he given the task of developing a centralized reference catalog for forestry literature.

(1) The Oxford classification would be used. It is the most satisfactory of existing systems, has international sanction, and its use avoids the necessity of reclassification of abstracts and bibliographic citations from the major bibliographic centers.

(2) All information would be put on cards, eliminating the use of envelopes for cut abstracts.

(3) The choice of card size poses a difficult question. Standard 3-by-5-inch cards are furnished by bibliographic centers, but are too small to contain abstracts without reduction. The present 4-by-6-inch cards will accommodate most abstracts without reduction, but folding of the longer abstracts on the back side is necessary. Larger 5-by-7-inch cards might prove most satisfactory in the long run.

(4) All abstracting and bibliographic services should be subscribed to—and all pertinent abstracts and bibliographic citations would be separated and either reproduced or pasted on the cards (if not supplied as cards of the proper size). *Forestry Abstracts* and *Biological Abstracts* would form the basis for the current

literature records. All available annotated bibliographies would be drawn upon. Simple bibliographic citations would be included in the absence of an abstract.

(5) Reproduction facilities would be developed for the printing of sufficient additional cards for cross-indexing.

(6) Facilities for photographic reduction and reproduction would be developed for the printing of entire short papers on the standard cards. For instance, the large number of one- and two-page "research note" type of articles should be capable of reduction to a single 5-by-7-inch card, thus permitting the inclusion of the entire work rather than of an abstract. In order to permit the simultaneous handling of a large number of cards dealing with a single subject, the writer is inclined to limit the amount of reduction so that the product can be read with the naked eye or under simple magnification.

By following the procedures outlined above, it should be possible to construct and maintain a central literature file for world forestry literature at a very moderate cost. This file would take the form of a single catalog containing abstracts, citations, and photographic reductions of brief articles. With adequate cross-referencing, mechanical retrieval systems would not be necessary. As time and facilities became available—and this would undoubtedly require international cooperation—citation cards could be expanded to abstract cards, and abstract cards in many instances could be expanded to more detailed record cards or even to complete reproduction cards. Personnel trained as scientists in the field should take an active part in the operation, rather than trusting entirely to professional abstractors or librarians.

The advantages of the system are its complete practicability, its low cost, and the fact that it becomes immediately usable once established. The prototype developed by the author over a six-year period is constantly used and has resulted in its users achieving a better comprehension of the literature relating to their problems in less time than previously required for a less satisfactory survey.

The Information-Gathering Habits of American Medical Scientists

SAUL HERNER

Since the Royal Society Scientific Information Conference in 1948, and before, there has been a growing interest in the methods by which scientists obtain information and communicate information to one another. For the most part, this interest has stemmed from a rapid growth in the world's scientific activity and a corresponding growth in the written and published output of scientists. It has also arisen from an increased appreciation of the economic and political significance of scientific information.

One manifestation of the increase in the appreciation of the political and economic aspects of scientific information has been the organization of efforts, in the United States and abroad, to increase the availability of Soviet scientific information. This preoccupation with Soviet information came to a head with the launching of the first Russian satellite.

But even before the advent of the satellite there were various Soviet information programs in operation among a number of agencies of the United States Government. These programs took various forms and went in various directions, but all shared the common goal of making Soviet scientific information more readily available to the American scientist.

One of these Government Soviet information programs gave rise to the study upon which the present paper is based. The specific mission of this program was the dissemination of information on Soviet medical research to American medical scientists. In order to ascertain the most effective directions that such a program might take, a project was organized to determine, on a statistically dependable basis, the current use that 500 American medical scientists in 59 institutions and organizations make of Soviet information in their fields.

For purposes of comparison, the project was designed to consider not only the scientists' use of Soviet information, but also their use of information in general and foreign-language information in general. Thus, the project under

SAUL HERNER Herner and Company, Washington, D.C.

discussion was actually a trichotomous study of the habits and patterns of American medical scientists in the use of information in general, foreign-language information in general, and Soviet information in particular.

The thinking behind this project design was that the most expeditious way to make useful Soviet information available to American medical scientists was by utilizing the already established channels and mechanisms by which they obtain other types of information. Obviously, if Soviet information is channeled to American scientists through uncommon or unfamiliar media, it will not be used as effectively as it would be if it were channeled to them through media that they are already using. Thus, the purpose of the project was to determine the media used by the respondent-scientists to obtain Soviet information at present and in the very recent past. Where respondents had made no recent use of Soviet information, the study sought to determine how they gained access to other foreign-language information. In cases where respondents had not made recent use of any foreign-language information, the study turned to their use of information in general. The focus always was on existing, familiar channels of information.

The findings of the respondents' use of Soviet and other foreign-language information have already been disseminated in the form of a report entitled, “The Use of Soviet Medical Research Information by American Medical Scientists,” which was prepared and distributed in 1957. The present paper treats of the findings of the part of the study dealing with their use of information in general.

METHOD OF STUDY

The study was conducted by means of detailed interviews with the 500 respondent-scientists. The interviews were conducted by trained interviewers, all of whom had had experience in face-to-face interviews with working scientists. The interviewers were guided by carefully designed and pretested questionnaires. While the majority of the questions in the questionnaire were of the preceded or check-off type, they were kept sufficiently open-ended to permit answers other than those anticipated. The interviewers were instructed to record verbatim all information volunteered by respondents in answer to questions. In the case of discussion questions, the interviewers probed to obtain the most detailed answers possible. The average interview consumed approximately an hour and a quarter.

The 500 respondent scientists were selected randomly in 59 medical research institutions and organizations in New York, Philadelphia, Baltimore, Washington, Chicago, and Cleveland. These scientists represented all the major fields of medical research. Their median age was 36. Forty-five per cent had M.D. degrees; 45 per cent had Ph.D. or D.Sc. degrees; and the remaining 10 per cent had various other degrees.

The specific establishments from which the respondents were drawn were obtained from such reference works as the National Research Council's *Industrial Research Laboratories in the United States and Canada,* the American Council on Education's *American Colleges and Universities,* and the National Science Foundation's *Grants and Contracts for Unclassified Research in the Life Sciences.*

The selection of organizations for study was intentionally skewed so as to ensure the inclusion of the highest possible proportion of establishments having well-rounded library and information facilities. This was done to maximize the number of respondents likely to have had recent or current experience in the use of Soviet information. In view of this intentional bias, the sample cannot be considered representative of all American medical research activities, and the conclusions and implications drawn from the study are valid only for those organizations represented in the sample. However, in view of the size of the sample and the method of sampling, there is reasonable assurance of a tolerably small sampling error and a proportionately representative profile of the limited universe.

The method of study, the face-to-face interview, is based on well-established and generally recognized procedures developed and used for many years in polling and sampling survey activities. Various criticisms have been leveled against the use of face-to-face interviews for the purpose for which they were used in the study under discussion. These criticisms have come primarily from librarians and documentalists, and not from professionals in the survey field.

Actually, the specific question of the applicability of face-to-face interviewing techniques has been studied in considerable detail at the Survey Research Center of the University of Michigan. The results of the Michigan study established (if indeed there ever was any doubt about it) that "…even such a complicated matter as science can be handled using survey interview techniques. Meaningful answers can be obtained by this means to a series of searching questions…." (Cf. University of Michigan Survey Research Institute, "Science Writing and the Public; A Report of a Pilot Study for the National Association of Science Writers," p. 48, The Institute, Ann Arbor, 1955.)

METHODS OF KEEPING ABREAST OF CURRENT DEVELOPMENTS

The first question put to the respondents was the following: "How do you generally keep abreast of current scientific developments in your field?" This question produced a total of 65 different tools and techniques. However, only eight were named by 10 or more respondents.

Three tools or techniques stood out among all others. These were the regular scanning of research journals, attendance at meetings and lectures, and face-to-face contacts with colleagues. These three tools or techniques constituted 77 per cent of all those mentioned. The remaining five of the eight primary tools or techniques mentioned were indexing and abstracting publications, textbooks, review papers, correspondence with colleagues, and visits to other research organizations.

SOLVING PROBLEMS AND ANSWERING QUESTIONS

A second question (actually a set of questions) addressed itself to the methods used by the respondents to solve problems and obtain answers to questions that they could not answer from their own immediate knowledge. In order to minimize problems of recall and to obtain the most accurate responses possible, each respondent was first asked to describe a recent instance in which he had a problem or question that he was not able to answer from his own immediate knowledge or background. Then, focusing on the case described, the respondent was asked how he had gone about finding an answer or solution.

In all, a total of 12 major categories of problems and questions, and seven primary sources of solutions or answers were named. The categories of problems and questions are correlated with the sources of answers or solutions in Appendix I. It will be noted that the total number of times that the various sources of solutions and answers are mentioned in Appendix I is far in excess of 500, the number of respondents interviewed. This results from the fact that many of the interviewees used more than one means to obtain their answers or solutions.

Of the 12 major categories of problems and questions, the largest number were concerned with characteristics, occurrence, diagnosis, or treatment of specific diseases. The second, third, fourth, and fifth most prominent categories of problems or questions were: physiological or chemical makeup of tissues, organs, or organisms; analytical chemical methods; biological effects of chemicals, drugs, or radiations; apparatus, equipment, or supplies; and surgical or dissection techniques. The last two categories were mentioned with equal frequency.

As might have been expected, personal advice from colleagues constituted the most frequently used means of obtaining answers to questions and solutions to problems. Close behind were papers in scientific and technical journals. Indexing and abstracting publications were a rather distant third, followed by the utilization of cited references in literature read, and the conduct of laboratory experiments.

SOURCES OF IDEAS

Still another general category of use to which information might be put was investigated by means of the question, "Do you recall where you got the idea (or inspiration) for your present or most recent project?" Here again, the question was worded in such a way as to minimize the problem of recall, and to give the respondent a specific case upon which to focus.

In the greatest number of cases, it developed that the respondent's current or most recent project was merely a continuation or out-branching of work he had been doing. Aside from this self-generating source of ideas, the primary source was face-to-face communications and discussions with colleagues. Then came reading the literature, observations made in the course of treating patients, assignments or suggestions from superiors, hearing lectures, and various others. There were in all 11 major sources of ideas among the respondents studied. These are listed in Appendix II.

METHODS OR TOOLS FOR LOCATING OR BECOMING AWARE OF SOURCES OF INFORMATION

The question of how the respondents learn of the existence of or locate publications or other sources of information which might be useful was approached by handing each respondent a list of common bibliographical tools or techniques and asking them to state which they had used in the previous six months. "Chance or accident" was included in the list of tools or techniques.

The relative use of the various tools and techniques is given in Appendix III. In general, there was significant use made of all those listed, and the intervals among the percentages of respondents using the various tools and techniques were relatively small. There were no significant differences in the percentages of the respondents who had used the three most frequently used tools or techniques. These three most prominent tools or techniques were the following: footnotes or other cited references, chance or accident, and indexing and abstracting publications. After these came personal recommendations, personal reference files, book reviews, library card catalogs, publishers' advertisements, library acquisitions lists, and separate bibliographies, in that order. The average respondent had used an average of 7.6 of the aforementioned tools or techniques in the previous six months.

TOOLS AND METHODS USED IN LITERATURE SEARCHES

As an enlargement of the question on how the respondents locate or become aware of useful sources of information, each was asked to describe the most recent question or problem that involved a literature search and to tell how he went about doing the search. Once again, the question (or set of questions) was so worded as to induce the respondent to focus on a specific, recent case, thereby minimizing the problem of recall.

The types of problems described paralleled approximately those described in answer to the set of questions on how the respondents get answers or solutions to questions or problems that they cannot handle from their own immediate knowledge or background. The one notable difference was that a fairly large number of the respondents had conducted literature searches not to solve a specific problem or to obtain the answer to a specific question, but to obtain background information in connection with the writing of a paper or book.

There was a total of 75 different tools or techniques used by the respondents in the conduct of their searches. Of these, 12 were named by 10 or more respondents. Of these 12, three were distinctly more prominent than the others. These three were: going directly through the most likely journals and browsing, consulting indexing and abstracting publications, and following up footnotes and cited references. The role of personal recommendations was significant in the conduct of formal literature searches, but not nearly as significant as in the respondents' other information-seeking activities. The relative significance of the 12 major searching tools or techniques is given in Appendix IV.

SIGNIFICANCE OF LITERATURE SEARCHES

One final aspect of literature searching which was considered in the present study was the significance of searches as a means of avoiding unnecessary duplication of effort. Each of the respondents was asked two questions in this connection. The first was: “When you started your most recent research project, did you precede it with a literature search to see what had already been done on the problem?” The second question was: “Would you say that you generally precede or begin new research studies with reviews of the literature to see what has already been done in the fields in which you're going to do your research?”

In answer to the first question, 303 of the respondents had preceded their most recent projects with literature searches, and 197 had not. In answer to the second question, 341 of the respondents stated that they did, as a rule, make a literature search before starting a new project.

While no further questions were asked about preceding projects with searches, all qualifying remarks made along with the “Yes” or “No” answers were recorded, and detailed analyses were made of these remarks. Particular attention was given to the qualifying remarks of those respondents who answered either or both of the questions in the negative.

From the analysis of remarks, it developed that many of the respondents who answered negatively did do searches in connection with their projects, but they did them in the *course* of the projects rather than *before* they started. Another common qualification of the negative answers was that most new projects are not preceded by searches because they are not really new projects but continuations or outgrowths of old projects. Here again, we have an instance of literature searching on a continuing, informal basis, rather than a formal basis.

Of the various reasons for not searching the literature, the following cropped up most frequently: Respondent had previous knowledge of field and its literature; there was a lack of literature in the field; literature searches bias or inhibit a research worker; literature frequently contains errors; literature searches are not the job of the respondents.

CONCLUSIONS

The primary conclusion that can be drawn from the foregoing paragraphs is a reaffirmation of the significant role of personal contacts in the getting and transmitting of scientific and technical information. Face-to-face communications emerged as the most important method of answering questions or solving problems, the most important source of ideas (after the respondents' own researches and observations), and as one of three most important means of keeping abreast of current developments.

While the role of personal contacts was still quite formidable in the conduct of literature searches and in locating or becoming aware of useful sources of information, footnotes and cited references and indexing and abstracting publications were significantly more important.

The obvious implication of these findings is that in the day-to-day aspects of information communication, informal tools are foremost, but where formal searches are done the searcher leans more heavily on tools of a more scholarly or formally bibliographic nature.

APPENDIX I CATEGORIES OF PROBLEMS AND QUESTIONS CORRELATED WITH SOURCES OF SOLUTIONS OR ANSWERS

	Methods of solution						
Problems or questions	*Texts or Monographs*	*Cited refs.*	*Personal contacts*	*Indexing and Abstract publications*	*Journals*	*Own experiments*	*Miscellaneous*
Analyt. chemical methods	11	4	36	11	21	16	11
Structure or synthesis of compounds	0	6	0	6	14	0	24
Physical or chemical properties of compounds or elements	0	5	14	14	12	11	0
Apparatus equipment or supplies	12	5	14	0	0	0	0
Biological effects of drugs, chemicals, or radiations	0	0	54	16	4	9	11
Pathological activities, physiology, or determination of microorganisms	11	0	12	0	11	16	6
Physiological properties or chemical makeup of tissues, organs, or organisms	18	16	10	4	9	12	0
Physiological analytical techniques	6	0	14	3	4	4	21
Anatomy or embryology of tissues or organs	12	0	0	5	14	0	6
Surgical or dissection techniques	12	5	4	6	13	5	0
Information on specific diseases	36	48	25	41	60	16	15
Experimental methodology	0	0	6	4	0	0	0
Miscellaneous	0	0	16	19	31	0	11
Totals	118	89	205	129	193	89	115

APPENDIX II SOURCES OF IDEAS FOR CURRENT PROJECT

Source	*Number of respondents*
Own previous work	215
Colleagues	133
Reading literature	105
Observation of patients	70
Assignments or suggestions from superiors	58
Hearing lectures	40
Omissions in the literature	38
Disagreement with literature	18
Teaching activities	13
Taking courses	8
Manufacturers or suppliers	8
Miscellaneous	53

APPENDIX III TOOLS OR METHODS USED FOR LOCATING OR BECOMING AWARE OF SOURCES OF INFORMATION

Tools or methods	*Number of respondents*
Footnotes or other cited references	486
By chance or accident, while looking through publications	481
Indexing and abstracting publications	477
Personal recommendations	439
Personal reference file	405
Book reviews	351
Library card catalogs	346
Publishers' advertisements	332
Library acquisitions lists	291
Separate bibliographies	212
Others	12

APPENDIX IV PRIMARY TOOLS OR TECHNIQUES USED IN LITERATURE SEARCHES

Tools or techniques	*Number of respondents*
Consulted colleagues	215
Correspondence	35
Footnotes and cited references	258
Texts and monographs	168
Indexes and abstracts	263
Went to journals	348
Journal indexes	33
Asked librarians	48
Card catalogs	25
Personal reference file	113
Review papers	123
Bibliographies	33

Use of Scientific Periodicals

D.J.URQUHART

In the United Kingdom it has been decided to plan a National Lending Library for Science and Technology, and it is intended that this library should take over from the Science Museum Library the general responsibility for providing a lending service to organizations in the United Kingdom. To obtain some data for this planning operation it was decided to analyse the issue records of serial publications of the Science Museum Library, which at present has the largest collection of scientific literature in the United Kingdom. The Science Museum Library is actually within the buildings of the Imperial College of Science and Technology (South Kensington, London, S.W.7), and functions, in practice, as the main library of that college. The library has a reading room which is open to the public, and it lends to the staff of the Science Museum and the Imperial College, and to over 1200 outside organizations. These include universities and colleges, industrial organizations, research organizations, and government departments.

At present the library is receiving about 10,000 current serials. In all, it contains some 430,000 volumes. The subjects covered include all the sub-divisions of the pure and applied sciences, with some restrictions in the clinical medicine field. About 80% of the total volume of receipts and issues are serial publications.

The records examined were the counterfoils of issue forms for literature returned to the shelves from the reading room or from borrowers during 1956. An unknown number of these forms may have been lost, but the missing forms would probably amount only to a few per cent., and they are likely to have been a random selection of the total.

The primary object of the analysis was to provide data for planning purposes. It was hoped, for instance, that the data would answer such questions as *(a)* which serials are so heavily used that they must be held by the National Lending Library, *(b)* what back runs of serials should be collected, *(c)* which serials should be duplicated or bound in parts, *(d)* which serials are so little used that one copy either in the future Science Museum Library or the National

D.J.URQUHART Lending Library Unit, Department of Scientific and Industrial Research, London, England.

Lending Library should be sufficient. However, as this survey of actual use of scientific serials is probably the largest which has ever been made, it was thought that a general account of the results would be of interest.

The analysis was made by punching a Hollerith card for each item issued from the shelves. This gave *(a)* the shelf mark of the serial (this is unaffected by a change of title); this was added to the counterfoils where necessary; *(b)* the type of user: A, borrower in the Science Museum or Imperial College (i.e., "internal loans"), B, other borrowers (i.e., "external loans"), C, reader in the Science Museum Library's reading room; (c) the date of publication of the item used where given. In all, 87,255 cards were punched, and they were divided up as follows:

A	Loans to Science Museum and Imperial College	2,255
B	Loans to other borrowers	53,216
C	Issues in the reading room	27,161

4623 (5.3%) of the forms were rejected, mainly because they did not give the publication date of the item issued. The remainder were sorted by date, and this gave the results shown in Table I.

TABLE I

Publication date	A *Internal loans*	B *External loans*	C *Reading room issues*	*Total*
1857–1899	50	665	420	1,135
1900–1909	39	562	412	1,013
1910–1919	43	801	568	1,412
1920–1929	177	2,073	1,474	3,724
1930–1939	417	6,284	4,104	10,805
1940–1949	459	10,013	6,596	17,068
1950–1954	600	14,518	8,723	23,841
1955–1956	470	18,300	4,864	23,634
Total	2,255	53,216	27,161[a]	82,632

[a] Owing to the methods by which the forms were selected two copies of the same reading room issue form may have been punched on some occasions. This has not happened in more than 10% of cases, and will not have affected the total number of periodicals used.

In interpreting these results the following should be noted:

1. The Science Museum Library lends in the main to organizations which have some library resources of their own.
2. The current volumes of some 500 periodicals are available on open access in the reading room.
3. The Science Museum Library does not lend reference publications such as abstract publications, and during 1956 the 1940 onwards volumes of 72 periodicals were on a non-loanable list. This list is given in Appendix A.
4. Binding operations during 1956 will have reduced the availability of material published during 1954.

5. Whilst there is no Imperial College Library several of the departments of the college have their own libraries.

The analysis by serial shelf marks showed that, in all, 7064 serials were used.

The following details of the number of serials used are available:

Number of serials borrowed by the local staffs	1,063
Number of serials borrowed by others	5,632
Number of serials used in the reading room	3,518
Total number of serials used by all types of users where publication date was 1950 or afterwards	4,753
Total number of serials used by all types of users where publication date was 1949 or earlier	4,058

From the general point of view the most interesting figures relate to the loan issues to external organizations. Tables II to V illustrate the results obtained and make it possible to compare the number of issues from the Science Museum Library with the total number of holdings as given by the British Union Catalogue of Periodicals (BUCOP) for the main libraries in the United Kingdom.

TABLE II The 10 titles most frequently borrowed by external organizations

Issues	*Title*[a]	*Current holdings in BUCOP*
382	*Proceedings, Royal Society, A* (Vol. 1, 1832–)	71
250	*Journal of Physical Chemistry* (Vol. 1, 1896–)	42
244	*Philosophical Magazine* (1789–)	54
240	*Science* (U. S.) (1883–)	55
223	Declassified documents, U.S. Atomic Energy Commission (1946–)	—
223	*Proceedings, Institution of Electrical Engineers* (1872–)	51
200	*Product Engineering* (Vol. 2, 1931–)	11
198	*Biochemical Journal* (Vol. 1, 1906–)	81
184	*Journal, Chemical Society* (1849–)	86
181	*Journal, Institution of Mechanical Engineers* (1939–)	62

[a] Science Museum Library holdings are given in parentheses.

TABLE III A sample of 10 titles each borrowed 20 times by external organizations

Title[a]	*Current holdings in BUCOP*
Proceedings, Institution of Civil Engineers (Vol. 1, 1952–)	52
Journal of Applied Mechanics (Vol. 2, 1935–)	26
Practitioner (Vol. 74, 1905–)	33
Archiv für Protistenkunde (Vol. 1, 1902–)	16
Rubber Journal (Vol. 68, 1924–)	14
Process (formerly Photogram) (Vols. 1–10, 46–)	9
Siemens-Zeitschrift (Jahrg. 9–, 1929–)	2
Bulletin, Research Council of Israel (Vol. 1, 1951–)	8
Annals of Human Genetics (Vol. 1, 1925–)	22
Bacteriological Reviews (Vol. 1, 1937–)	42

[a] Science Museum Library holdings are given in parentheses.

TABLE IV A sample of 10 titles each borrowed twice by external organizations

Title[a]	*Current holdings in BUCOP*
American Journal of Clinical Nutrition (Vol. 3, 1955–)	—
Elektrowärme-Technik (Jahrg 6–; 1955–)	1
Medicine Monographs (1925 and 1930)	1
Television (1939, 1950–)	1
University of California publications in child development (Vol. 1, 1949–)	2
Proceedings, Cotteswold Naturalists Field Club (Vol. 12, 1896–)	15
Indian Journal of Radiology (Vol. 3, 1949–)	2
Indian Journal of Genetics and Plant Breeding (Vol. 1, 1941–)	6
Philippine Journal of Agriculture (Vol. 1, 1930–)	8
Journal, Mechanical Laboratory, Japan (European Language Edition) (Vol. 1, 1955–)	—

[a] Science Museum Library holdings are given in parentheses.

TABLE V A sample of 10 titles not used by external organizations

Title[a]	*Current holdings in BUCOP*
Acta Astronomica (Warzawa—Series C, Vol. 5, 1953–)	4
Aquila Budapest (1929–38, 1943–)	5
Boletin, Casa do Dovro (Vol. 3, 1948–)	1
Bulletin Agricole-Haiti (Vol. 1, 1950–)	—
Hormone—Oss (Vol. 11, 1949–)	2
Journal des Observateurs (Vol. 1, 1915–)	—
Pig Farming (Vol. 1, 1953–)	1
Report, University of Washington, Engineering Experiment Station (Vol. 1, 1929–)	—
Revista Vinicola y de Agricultura (1935–36, 1948–)	1
Vestnik—Ceska Akademie ved a Umeni (Vol. 34, 1925–)	2

[a] Science Museum Library holdings are given in parentheses.

External organizations will naturally only borrow from the Science Museum Library scientific literature which they do not hold themselves, or which they cannot obtain from some more accessible collection. Thus the external loan demand on the library is, in general, only a residual demand, although many bodies may make a habit of always applying to this library first. Nevertheless, possibly because so many external organizations (some 1200) use the Science Museum Library, it appears from the examples in Tables II to V, and from Table VI, that the use of the copies of a serial in the library is a rough indication of its total use value in the United Kingdom. An analysis of the type of libraries which hold little used serials indicates that, in general, they are the general and not the special libraries. So that it is probable that, as a rule, the copy of the serial publication in the Science Museum Library is more used than a copy of the same serial elsewhere. However this may be, it is very-clear that it is not the abundance of sets of certain serials in other libraries which results in the Science Museum Library's holdings of these serials being so little used.

TABLE VI

Type of serial	*Average number of current sets in BUCOP*[a]
10 most frequently used (Table II)	57
Sample of 10 serials each used 20 times (Table III)	22.4
Sample of 10 serials each used twice (Table IV)	4.5
Sample of 10 serials not used (Table V)	2.3

[a] Only the titles located in BUCOP were used in obtaining average.

The number of titles, according to degree of use on external loans, is given in Table VII.

TABLE VII

Requisitions per title	*Titles*
100 or more	60
50–99	193
40–49	92
30–39	136
20–29	229
10–19	541
5`9	714
4	283
3	403
2	791
1	2,190

An analysis of the serial titles used showed that 2769 of them were not current according to the *Hand List of Short Titles of Current Periodicals in the Science Library, Eighth Edition, 1956.* This list contains references to 9120 serial publications, and 4821 of these were not used at all during 1956. The greater proportion of users would have been using, during 1956, the 1953 edition of the Hand List, which contains a smaller number of titles.

These figures do not necessarily mean that the Science Museum Library is collecting some 4821 serials which will never be used, but they do indicate that the demand for a large number of titles is very small. For instance, if the library contains 7500 serials, each of which is used on an average once every two years, we should expect to find in a particular year that of these, assuming random distribution of demand

Number of titles not used was	4,548
Number of titles used once was	2,274
Number of titles used twice was	568
Number of titles used thrice was	95

The total use of a periodical is not as useful a figure as the amount of use per year of publication. The number of titles for which the 1900–1909 volumes were borrowed by external organizations 10 or more times was 9. The number of titles which were found with the same minimum degree of use in different periods is shown in Table VIII.

TABLE VIII Number of titles used once or more per year of publication by external borrowers

Period	*Titles*
1900–1909	9
1910–1919	14
1920–1929	40
1930–1939	158
1940–1949	258
1950–1954	872

Appendix B gives a list of the titles which were borrowed 10 or more times for any of the publication decades 1900–1949, and indicates the decades concerned.

This list should be of value to special libraries in the United Kingdom in indicating what serial literature published in the period 1900–1949 they should consider obtaining or retaining.

The average period of loan of a serial publication lent by the Science Museum Library is about 23 calendar days. Hence, a volume which is lent on an average once per year should be available on demand for 94% of the time.

Considerations such as these make it possible to determine what must be done regarding binding periodicals in parts, or duplicating copies to achieve any particular standard of service (e.g., a minimum of 90% available on demand for any title). The survey of the actual use data indicates that, as a rule, binding parts of periodicals separately in a light binding would be cheaper in providing an improved standard of service for frequently used serials than duplicating or triplicating. These considerations are, of course, based on the assumption that for any title and time since publication the demand for it is random, but has a given average value. This assumption has not been tested. Theoretically it would be possible to test the assumption by comparing the theoretical number of times material is on loan when required with the number found in practice, but this would involve more data regarding the state of the bindings and of the volumes per year than is immediately available on the punched cards used.

CONCLUSION

The primary purpose of this paper was to record the facts as far as they are available, and to leave to others the application of the data to the problems which face the users of science literature throughout the world.

The analysis, however, confirms one general idea. Extensive use of scientific literature is confined to a small fraction of the total output. Even in a library which is designed to deal with the residual demand from libraries, about 1250 serials (or less than 10% of those available if the non-current serials are included) are sufficient to meet 80% of the demand for serial literature.

Indeed, the figures suggest that perhaps three-quarters of the current serials in the Science Museum Library are so little used that one loan copy of these serials somewhere in the United Kingdom should be sufficient to meet the needs of all users in the United Kingdom.

This analysis focuses attention on the old questions: (1) Is the small use of a large number of serials due to the low value of, or to ignorance about, their contents? (2) If it is ignorance, what can be done to eliminate it? (3) If the contents are designed only for a very small audience, should the papers continue to be published in the traditional way?

ACKNOWLEDGMENTS

The author wishes to acknowledge the assistance which he has received from Mr. H.T.Pledge, the Keeper of the Science Museum Library, and his staff, and from the staff of the Hollerith Section at the National Physical Laboratory. He also wishes to stress that he is personally responsible for the interpretations of the data given.

APPENDIX A SCIENCE MUSEUM LIBRARY: LIST OF NON-LOANABLE PERIODICALS

Acta Physicochimica U.R.S.S.
Analytical Chemistry
Angewandte Chemie
Annalen der Physik
Annales de physique
Annals, New York Academy of Sciences
Annual Reports on the Progress of Chemistry
Bell System Technical Journal
Berichte deutschen chemischen Gesellschaft
Bulletin, Académie des Sciences de l'U.R.S.S. (Izvestiya) (all series)
Bulletin, Société chimique de France
Canadian Journal of Botany
Canadian Journal of Chemistry
Canadian Journal of Medical Sciences
Canadian Journal of Physics
Canadian Journal of Research (all sections)
Canadian Journal of Technology
Canadian Journal of Zoology
Chemical and Engineering News
Chemical Engineering Progress
Chemie-Ingenieur-Technik
Chemische Berichte
Comptes rendus, académie des sciences (Paris)
Comptes rendus académie des sciences de l'*U.R.S. S.*
Doklady, Akademii Nauk SSSR
Electrical Engineering
Helvetica Chimica Acta
Helvetica Physica Acta
Industrial and Engineering Chemistry (all editions)
Iron Age
Journal, Acoustical Society of America
Journal, American Chemical Society
Journal, Electrochemical Society
Journal, Franklin Institute
Journal, Optical Society of America
Journal of Applied Mechanics
Journal of Applied Physics
Journal of Biological Chemistry
Journal of Chemical Physics
Journal of Metals
Journal of Petroleum Technology
Journal of Research, National Bureau of Standards (U.S.)
Mechanical Engineering
Metals Technology
Mining Engineering
Mining Technology
Petroleum Technology
Physica

Physical Review
Proceedings, Institute of Radio Engineers (New York)
Reports on Progress in Physics
Reports on the Progress of Applied Chemistry
Review of Scientific Instruments
Reviews of Modern Physics
Steel
Transactions, (American) Electrochemical Society
Transactions, American Institute of Chemical Engineers
Transactions, American Institute of Electrical Engineers
Transactions, American Institute of Mining and Metallurgical Engineers (all sections)
Transactions, American Society of Mechanical Engineers
Transactions, Institute of Radio Engineers (Professional groups) (New York)
V.D.I. Zeitschrift
Zeitschrift für analytische Chemie
Zeitschrift für anorganische und allgemeine Chemie
Zeitschrift für Physik
Zeitschrift für physikalische Chemie
Zhurnal analitichesko‘ khimii (J. of Analytical Chemistry)
Zhurnal eksperimental’ no‘ i teoretichesko‘ fiziki (J. of Experimental and Theoretical Physics USSR)
Zhurnal fizichesko‘ khimii (J. of Physical Chemistry USSR)
Zhurnal obshche‘ khimii (J. of General Chemistry USSR)
Zhurnal prikladno‘ khimii (J. of Applied Chemistry USSR)
Zhurnal tekhnichesko‘ fiziki (J. of Technical Physics USSR)

APPENDIX B SCIENCE MUSEUM LIBRARY: TITLES BORROWED TEN OR MORE TIMES PER DECADE, 1940–1949

Title	Years for which periodical was borrowed ten or more times per decade
Acta Chemica Scandinavica, 1947	1940–49
Acta Physiologica Scandinavica Supplementa, 1940	1940–49
Agronomy Journal, 1932	1940–49
Air Conditioning, Heating, and Ventilaing, 1929	1940–49
Aircraft Engineering, 1929	1930–49
American Dyestuff Reporter, 1925	1930–49
American Mineralogist, 1916	1930–39
American Naturalist, 1868	1940–49
American Journal of Botany, 1924	1940–49
American Journal of Hygiene, 1938	1940–49
American Journal of Public Health and the Nation's Health, 1932	1940–49
American Journal of Physics, 1933	1940–49
American Journal of Physiology, 1898	1930–49
American Journal of Roentgenology—Radium Therapy and Nuclear Medicine, 1930	1940–49
American Journal of Science, 1818	1910–49
American Journal of the Medical Sciences, 1936	1940–49
Analyst (London), 1882	1930–49
Analytica Chimica Acta, 1947	1940–49
Anatomical Record, 1908	1930–49
Annalen der Chemie, 1832	1900–39
Annalen der Physik, 1848	1900–09, 1920–39*
Annales de chimie, 1789	1940–49
Annals of Applied Biology, 1914	1920–39
Annals of Botany, 1887	1920–29
Annals of Tropical Medicine and Parasitology, 1907	1940–49
Annals of Mathematical Statistics, 1930	1930–39
Archiv für das Eisenhuttenwesen, 1928–44; 1949–	1930–49
Archiv für Mikrobiologie, 1930	1930–39
Archiv für Experimentelle Pathologie und Pharmakologie, 1938	1940–49
Archiv für Protistenkunde, 1902	1930–39
Archives of Biochemistry and Biophysics, 1942	1940–49
Arkiv för Kemi, 1903	1940–49
ASTM Bulletin, 1928	1940–49

*Indicates title appearing also in Appendix A.

Title	*Years for which periodical was borrowed ten or more times per decade*
Astrophysical Journal, 1895	1940–49
Audio Engineering, 1941	1940–49
Australasian Engineer, 1930	1940–49
Australian Journal of Experimental Biology and Medical Science, 1924	1940–49
Australian Journal of Science, 1938	1940–49
Automobile Engineer, 1924	1940–49
Bell System Technical Journal, 1925	1930–49
Biochemical Journal, 1906	1920–49
Biochemische Zeitschrift, 1906	1920–49
Biological Bulletin, Marine Biological Laboratory, 1900	1930–49
Biological Reviews, Cambridge Philosophical Society, 1925	1940–49
Biometrics, 1946	1940–49
Biometrika, 1902	1940–49
Botanical Gazette, 1875	1920–49
Botanical Review, 1935	1940–49
Botaniska Notiser, 1929	1940–49
British Journal of Experimental Pathology, 1929	1940–49
British Journal of Industrial Medicine, 1944	1940–49
British Journal of Radiology, 1926	1940–49
British Plastics, 1930	1940–49
Brown Boverie Review, 1927	1940–49
Bulletin, American Association of Petroleum Geologists, 1917	1940–49
Bulletin, American Mathematical Society, 1927	1940–49
Bulletin, American Physical Society, 1937	1940–49
Bulletin, Association Suisse des Electriciens, 1928	1940–49
Bulletin, Bureau of Mines (U.S.A.) 1912	1930–39
Bulletin, Geological Society of America, 1925	1940–49
Bulletin, Société de chimie biologique, 1914	1940–49
Bulletin, Société chimique de France, 1858	1930–39*
Bulletin of Mathematical Biophysics, 1939	1940–49
Canadian Chemical Processing, 1925	1940–49
Canadian Journal of Research, 1929	1930–39
Cereal Chemistry, 1932	1940–49
Chemical Engineering (New York), 1903	1930–49
Chemical Engineering Progress, 1908	1930–39*
Chemical Reviews, 1925	1930–49
Chemische Berichte, 1868–1943; 1947–	1900–09, 1920–29*
Chemische Weekblad, 1924	1940–49
Chemistry and Industry, 1882	1920–49
Chimie et industrie (Paris), 1918	1940–49
Civil Engineering (New York), 1931	1940–49
Cold Spring Harbor Symposia on Quantitative Biology, 1933	1940–49
Comptes rendus académie des sciences (Paris), 1835	1900–39*
Comptes rendus des séances, société de biologie, 1917	1930–49
Comptes rendus des travaux, laboratoire Carlsberg, 1878	1940–49
Contributions from Boyce Thompson Institute, 1929	1940–49
Current Science, 1932	1940–49
Discussions, Faraday Society, 1947	1940–49
Ecology, 1920	1930–39
Economic Geology, 1906	1940–49
Electrical Engineering, 1934	1930–39*
Electrical Journal, 1878	1930–39
Electrical Review, London, 1872	1940–49
Electronics, 1933	1930–39
Electronic Engineering, 1928	1940–49
Endocrinology, 1918	1940–49

*Indicates title appearing also in Appendix A.

Title	Years for which periodical was borrowed ten or more times per decade
Engineer, 1856	1930–49
Engineering, 1866	1910–19; 1930–49
Engineering News Record, 1922	1930–49
Escher Wyss News, 1928	1930–39
Experimentia, 1945	1940–49
Federation Proceedings (Federation of American Societies for Experimental Biology), 1942	1940–49
Food Industries, 1931	1940–49
Food Research, 1936	1940–49
Food Technology, 1947	1940–49
Forschung auf dem Gebiete des Ingenieurwesens, 1931–1940, 1943; 1949–	1930–39
Foundry Trade Journal, 1921	1940–49
Fuel, 1922	1940–49
General Electric Review, 1917	1920–49
Geological Magazine, 1864	1920–29
Glass Industry, 1928	1940–49
Growth, 1937	1940–49
Heating, Piping and Air Conditioning, 1929	1930–49
Helvetica Chimica Acta, 1918	1930–39*
Industrial and Engineering Chemistry, 1909	1920–39*
Industrial Chemist and Chemical Manufacturer, 1925	1940–49
Information Circular, U.S. Bureau of Mines, 1926	1940–49
Ingenieur Archiv, 1930–44; 1949–	1930–39
Instrument Practice, 1946	1940–49
Instruments and Automation, 1928	1940–49
Iron Age, 1922	1930–39
Iron and Coal Trades Review, 1900	1940–49*
Iron and Steel, 1928	1940–49
Iron and Steel Engineer, 1924	1940–49
Journal, Acoustical Society of America, 1929	1930–39*
Journal, American Ceramic Society, 1918	1930–49
Journal, American Chemical Society, 1879	1900–1939
Journal, American Concrete Institute, 1929	1930–49
Journal, American Leather Chemists Association, 1924	1940–49*
Journal, American Medical Association, 1930	1940–49
Journal, American Oil Chemists' Society, 1935	1940–49
Journal, American Pharmaceutical Association, Scientific Edition, 1929	1930–49
Journal, American Statistical Association, 1947	1940–49
Journal, American Water Works Association, 1924	1930–49
Journal, Association of Official Agricultural Chemists, 1920	1930–49
Journal, British Institution of Radio Engineers, 1939	1940–49
Journal, Chemical Society, 1849	1900–49
Journal, Council for Industrial and Scientific Research, Australia, 1927–48	1940–49
Journal, Franklin Institute, 1920	1930–39*
Journal, Indian Chemical Society, 1924	1930–49
Journal, Institute of Fuel, 1926	1930–49
Journal, Insitute of Metals, 1909	1940–49
Journal, Institute of Petroleum, 1915	1940–49
Journal, Institution of Heating and Ventilating Engineers, 1933	1930–49
Journal, Institution of Mechanical Engineers, 1939	1920–49
Journal, Iron and Steel Institute, 1871	1940–49
Journal, Oil and Colour Chemists' Association, 1918	1940–49
Journal, Optical Society of America, 1922	1930–39*
Journal, Royal Aeronautical Society, 1897	1940–49
Journal, Royal Statistical Society Series B, Methodological, 1936	1940–49
Journal, Scientific Research Institute (Tokyo), 1928	1940–49

*Indicates title appearing also in Appendix A.

Title	*Years for which periodical was borrowed ten or more times per decade*
Journal, Society of Dyers and Colourists, 1885	1940–49
Journal, Society of Glass Technology, 1917	1930–49
Journal, Society of Motion Picture and Television Engineers, 1916	1930–49
Journal, Textile Institute, 1910	1940–49
Journal, Washington Academy of Sciences, 1911	1940–49
Journal of Agricultural Science, 1905	1930–49
Journal of Applied Mechanics, 1935	1930–39
Journal of Applied Physics, 1931	1930–39*
Journal of Bacteriology, 1924	1930–49
Journal of Biological Chemistry, 1905	1910–39*
Journal of Cellular and Comparative Physiology, 1932	1930–49
Journal of Chemical Education, 1924	1930–49
Journal of Chemical Physics, 1933	1930–39*
Journal de chimie physique et de physico-chimie biologique, 1903	1940–49
Journal of Colloid Science, 1946	1940–49
Journal of Comparative Neurology, 1891	1930–49
Journal of Dairy Research, 1932	1930–49
Journal of Dairy Science, 1929	1940–49
Journal of Economic Entomology, 1908	1940–49
Journal of Experimental Biology, 1923	1940–49
Journal of Experimental Zoology, 1904	1910–19, 1930–39
Journal of General Physiology, 1919	1920–29; 1940–49
Journal of Geology, 1893	1930–49
Journal of Hygiene, 1901	1930–39
Journal of Immunology, 1940	1940–49
Journal of Industrial Hygiene, 1929–49	1930–49
Journal of Laboratory and Clinical Medicine, 1936	1940–49
Journal of Organic Chemistry, 1937	1930–49
Journal of Mathematics and Physics, 1922	1930–49
Journal of Pathology and Bacteriology, 1893	1930–39
Journal of Pharmacology and Experimental Therapeutics, 1936	1930–49
Journal of Physical Chemistry, 1896	1920–49
Journal of Physiology, 1878	1920–49
Journal of Polymer Science, 1946	1940–49
Journal of Research, National Bureau of Standards (U.S.A.), 1929	1920–39*
Journal of Nutrition, 1929	1930–49
Journal of Scientific and Industrial Research, 1942	1940–49
Journal of Scientific Instruments, 1924	1930–49
Journal of Sedimentary Petrology, 1932	1940–49
Journal of the Aeronautical Sciences, 1934	1940–49
Kolloid Zeitschrift, 1906	1930–49
Light Metals, 1938	1940–49
Lubrication Engineering, 1935	1940–49
Machine Design, 1932	1940–49
Machinery (London), 1918	1940–49
Machinery (New York), 1925	1940–49
Materials and Methods, 1930	1930–49
Mechanical Engineering, 1920	1930–39*
Mechanical Handling, 1891	1940–49
Mechanical World and Engineering Record, 1881	1940–49
Meddelanden, Sveriges Kemiska Industrikontor, 1931	1940–49
Metal Finishing, 1927	1940–49
Metal Industry (London), 1911	1940–49
Metal Progress, 1930	1940–49
Metal Working Production, 1900	1940–49
Metallurgia, 1929	1940–49

*Indicates title appearing also in Appendix A.

Title	*Years for which periodical was borrowed ten or more times per decade*
Mikrochemie, 1938–53	1930–49
Modern Plastics, 1932	1940–49
Mycologia, 1909	1940–49
Nature (London), 1870	1930–49
Naturwissenschaften, 1927–44; 1948–	1930–49
New England Journal of Medicine, 1938	1940–49
New Zealand Journal of Science and Technology, 1918	1940–49
Non-Destructive Testing, 1943	1940–49
Nucleonics, 1947	1940–49
Oil and Gas Journal, 1926	1930–49
Paint Manufacture, 1931	1940–49
Paint, Oil and Chemical Review, 1934	1940–49
Paper Industry, 1932	1940–49
Paper Trade Journal, 1924	1930–49
Petroleum Engineer, 1932	1940–49
Petroleum Refiner, 1925	1940–49
Pharmaceutica Acta Helvetiae, 1936	1940–49
Philips Research Reports, 1945	1940–49
Philips Technical Review, 1936	1940–49
Philosophical Magazine, 1789	1900–49
Philosophical Transactions, Royal Society A, 1667	1900–19, 1930–49
Philosophy of Science, 1934	1940–49
Photographic Journal, 1854	1940–49
Physica, 1933	1930–39*
Physical Review, 1894	1910–39*
Physikalische Zeitschrift, 1899–1945	1930–39
Physiological Reviews, 1933	1940–49
Phytopathology, 1911	1910–49
Plant Physiology, 1927	1930–49
Planta, 1926	1930–39
Plating, 1934	1940–49
Poultry Science, 1933	1940–49
Power, 1892	1940–49
Power Engineering, 1936	1940–49
Proceedings, American Academy of Arts and Sciences, 1846	1920–29
Proceedings, American Electroplaters Society, 1947	1940–49
Proceedings, American Gas Association, 1923	1940–49
Proceedings, American Philosophical Society, 1840	1940–49
Proceedings, American Society for Horticultural Science, 1949	1940–49
Proceedings, American Society for Testing Materials, 1923	1930–49
Proceedings, Cambridge Philosophical Society, 1925	1930–49
Proceedings, Geologists Association, 1865	1940–49
Proceedings, Helminthological Society of Washington, 1934	1940–49
Proceedings, Highway Research Board, U.S.A., 1929	1940–49
Proceedings, Indian Academy of Sciences, 1935	1930–49
Proceedings, Institute of Radio Engineers (New York), 1913	1930–39*
Proceedings, Institution of Civil Engineers, 1922	1940–49
Proceedings, Institution of Electrical Engineers, 1872	1930–49
Proceedings, London Mathematical Society, 1866	1930–39
Proceedings, National Academy of Sciences, U.S., 1915	1920–49
Proceedings, National Electronics Conference, 1945	1940–49
Proceedings, Physical Society, 1876	1910–49
Proceedings, Royal Academy of Sciences, Amsterdam, 1899	1930–39
Proceedings, Royal Society A, 1832	1910–49
Proceedings, Royal Society B, 1832	1920–49
Proceedings, Society for Experimental Biology and Medicine, 1929	1930–49

*Indicates title appearing also in Appendix A.

Title	*Years for which periodical was borrowed ten or more times per decade*
Proceedings, Society for Experimental Stress Analysis, 1943	1940–49
Proceedings, Soil Science, Society of America, 1937	1940–49
Product Engineering, 1931	1940–49
Products Finishing (Cincinnati), 1937–42; 1945–	1940–49
Protoplasma, 1927	1930–39
Public Health Reports, U.S., 1924	1940–49
Pulp and Paper Magazine of Canada, 1936	1940–49
Quarterly Journal of Mechanics and Applied Mathematics, 1948	1940–49
Quarterly of Applied Mathematics, 1943	1940–49
Quarterly Journal of Microscopical Science, 1853	1940–49
Quarterly Journal, Royal Meteorological Society, 1873	1940–49
R.C.A. Review, 1937	1940–49
Radio Engineering, 1930–49	1930–49
Radiology, 1933	1940–49
Receuil des travaux chimiques des Pays Bas, 1882	1930–49
Refrigerating Engineering, 1939	1930–49
Report of Investigations, U.S. Bureau of Mines, 1920	1930–49
Reports and Memoranda, Aeronautical Research Council, 1909–	1920–49
Review of Scientific Instruments, 1930	1930–39*
Reviews of Modern Physics, 1929	1930–39*
Revue Scientifique, 1863	1940–49
Rock Products, 1931	1940–49
Rubber Chemistry and Technology, 1932	1940–49
SAE Journal, 1924	1930–49
Science (U.S.), 1883	1920–49
Soap and Chemical Specialities, 1932	1940–49
Soil Science, 1916	1920–49
Stain Technology, 1927	1940–49
Steel Processing, 1926	1940–49
Structural Engineer, 1930	1940–49
TAPPI, 1920	1940–49
Transactions, American Electrochemical Society, 1909–49	1930–39*
Transactions, American Geophysical Union, 1922	1940–49
Transactions, American Institute of Electrical Engineers, 1885–1951	1920–39
Transactions, American Institute of Mining Engineers, 1873–1949	1930–39
Transactions, American Microscopical Society, 1924	1930–39
Transactions, American Society for Metals, 1925	1930–49
Transactions, American Society of Civil Engineers, 1897	1940–49
Transactions, American Society of Mechanical Engineers, 1897	1930–39*
Transactions, British Ceramic Society, 1902	1940–49
Transactions, British Mycological Society, 1897	1940–49
Transactions, Faraday Society, 1905	1920–49
Transactions, Illuminating Engineering Society, 1936	1940–49
Transactions, Institute of Metal Finishing, 1926	1940–49
Transactions, Institution of Chemical Engineers, 1923	1930–49
Transactions, Institution of Gas Engineers, 1903	1930–39
Transactions, Japan Society of Mechanical Engineers, 1935–40; 1947–	1930–39
Transactions, North East Coast Institution of Engineers and Shipbuilders, 1885	1940–49
Transactions, Royal Society of Canada, 3rd Series, 1924	1930–39
Textile Research Journal, 1932	1940–49
VDI Forschungsheft, 1931–1940; 1949–	1930–39
Welding Journal (New York), 1922	1940–49
Wireless Engineer, 1923	1930–49
Wireless World, 1914	1930–49
Zeitschrift für Analytische Chemie, 1862	1930–39*
Zeitschrift für Angewandte Mathematik und Mechanik, 1929–1943; 1948–	1930–39

*Indicates title appearing also in Appendix A.

Title	*Years for which periodical was borrowed ten or more times per decade*
Zeitschrift für anorganische und allgemeine Chemie, 1892	1920–39*
Zeitschrift für Elektrochemie, 1894	1920–49
Zeitschrift für Kristallographie, 1877	1930–39*
Zeitschrift für Lebensmittel-Untersuchung und -Forschung, 1898	1930–39
Zeitschrift für Metallkunde, 1911	1940–49
Zeitschrift für mikroskopisch-anatomische Forschung, 1924–39; 1950–	1930–39
Zeitschrift für Morphologie und Okologie der Tiere, 1924	1930–39
Zeitschrift für Naturforschung, A Astrophysik, Physik, physikalische Chemie, 1946	1940–49
Zeitschrift für Parasitenkunde, 1929–1939; 1949–	1930–39
Zeitschrift für Physik, 1920	1920–39*
Zeitschrift für physikalische Chemie, 1887–1944; 1950–	1900–09, 1920–39*
Zeitschrift für physiologische Chemie, 1877	1930–39
Zeitschrift für vergleichende Physiologie, 1924	1930–39
Zeitschrift für Zellforschung und mikroskopische Anatomie, 1924–44; 1949–	1930–39

*Indicates title appearing also in Appendix A.

SUMMARY OF DISCUSSION

DR. PHILIP M.MORSE, Panel Chairman, opened the discussion by stating that this first panel session concerning the informational needs of scientists in a way would set the stage for discussions of methods of meeting these needs later in the week. The question of the use of scientific information as a tool for research would be the central theme in this discussion. The quantity of scientific information is immense and is expanding exponentially. Its transmission from source to user is multi-channeled and very uneven. It cannot be concentrated into one or even a few types of channels. The papers imply that the greater the distance between fields of specialization, the less the efficiency of communication channels. Interdisciplinary communication is an extremely important problem to solve.

Some of the papers seem to indicate that scientists prefer face-to-face communication, which has a number of advantages. Would it be possible in the more mechanical methods of communication to incorporate a method of direct questioning and to select that part of the answer that is of immediate interest, as can be done in face-to-face communication?

Much scientific information is perishable goods, not treasure to be stored indefinitely. Material should not be left in the general research library too long.

Dr. Morse proceeded to discuss the paper by Halbert and Ackoff which reports on an attempt to see whether the field of operations research can help in this immensely complicated process of transfer of information. It is a brave effort that needs to be continued, but there is a long way to go. Whether the measures and criteria used are the best will have to be decided as time goes on. A number of the results of the study are of considerable interest, particularly the finding that chemists spend most of their working or productive time communicating, a finding that strengthens the belief that scientific communication is an important field. Whether the results are typical, even for chemists, is perhaps a question the reader can decide for himself. One can inquire whether operations research really has much chance of helping in this field, that is, whether the "operations" and interactions between the "system" and the people using it are sufficiently amenable to statistical analysis that operations research has much chance of being successful.

Dr. John W.Tukey pointed out that in the operations research approach there seems to be only an attempt to see what scientists do and not to ask what they think about things —and this in a process that is full of people.

Dr. Morse replied that he thought Dr. Tukey would agree that most of the so-called market research is really beside the point. One must take people's opinions about things with a grain of salt. Until now there has not been enough effort to find out what people do and too much of asking them what they think and want. Actually we should not have either type of study without the other.

Dr. Russell L.Ackoff commented on the question as to whether the results of the studies reported in the Conference papers are generalizable. In the Case Institute study, the population was well defined and the sample was selected in a precise way so that it would be possible to make some generalizations about the population of 55,000 chemists. He also stated that numerous examples can be given to illustrate that observational techniques obtain much different and more reliable data than questions asked of individuals about their behavior. It does seem to be important to get methods of observation that are subject to measures of reliability. As for the question of interdisciplinary communication, a profile of the chemist can be drawn from the results of the Case study that indicates that as he becomes a minority in a group he resorts more to written communication.

Dr. Mortimer Taube pointed out that the paper on the Case study stated that within the limits of the study it was not possible to develop a measure of effectiveness for scientific research. Instead the criterion of the availability of more time for research was used. If time is the only consideration, the best way to give scientists more time for research would be to abolish publications and information services so that they could spend all their time in the laboratory, but would science then be really productive? He suggested that there is confusion in the paper between the notions of productive science and scientific productivity.

Mr. Michael H.Halbert replied that the Case paper was not meant to convey the impression that scientific productivity is time spent in the laboratory. Scientific productivity is made up of the set of activities categorized in the study, including scientific communication, equipment set-up and use, data treatment, thinking and planning alone, etc. The scientist has only 100 per cent of his time to allocate. If one can find those factors that affect the way a scientist allocates his time, it might be possible to show that certain changes would make more time available for the scientist to use in whatever types of scientific activities he thinks best. It appears to be the consensus of chemists that it is most productive to allocate the largest proportion of their time to communication.

In speaking briefly of the Glass-Norwood study, Dr. Morse commented that many persons have a rather deep-seated suspicion of questionnaires concerning preferences. The Glass-Norwood study avoided this difficulty by asking scientists only about actions in specific cases. The results are of interest but only fragmentary.

Dr. Bentley Glass commented that the thing that came out of their little study that is most disconcerting and at the same time illuminating is the fact that the great majority of scientists questioned indicated that general conversation was the most frequent source of information about work that was essential to the development of their own ideas. He talked with a good many other scientists after the study was done and found that none of them disagreed with this feeling about sources of information. This finding implies that our existing sources of information, such as abstracting and indexing services, are not fully utilized, and it indicates a great degree of provincialism among American scientists. Dr. Glass said he cannot believe that the obtaining of essential information through casual conversation is a good means of obtaining knowledge of all that is going on in the world of science.

Dr. Chauncey D.Leake briefly discussed the Menzel paper pointing out that it clearly states the objectives of the interview survey at Columbia University. All channels of information exchange and all functions which scientific communication facilities are called upon to perform were considered. One point of interest is the significance of the unplanned mechanism, when a scientist while searching for a particular item stumbles upon something else useful. The question raised is "Are these cases merely individual accidents or do they occur with aggregate regularity?" Another question, that of motivation, which was mentioned by Dr. Menzel, merits further consideration. An important question to ask is "for what?" We so frequently take our objectives for granted that perhaps it is wise to take a critical look occasionally at our assumptions. Scientific information has manifold values, in logics, aesthetics, and ethics. Varying motivational factors are involved in producing, storing, and retrieving scientific information, but always the question is "for what?" If we can agree on some of the goals and objectives, then certain consequences are inevitable. These comments lead to a practical suggestion. Our modern concepts of science and democracy are derived from John Locke. One of his primary principles was that if persons in the community have access to the same body of reliable information, they will make wise decisions on policy matters. This principle was extended by Benjamin Franklin to the obligation of society to provide a free public library and by Jefferson to the provision of free public schools. Is it not time to extend the principle further to the obligation of society to provide for free distribution of scientific information at least to public and university libraries? Actually such distribution may become essential if the free flow of scientific information is to continue.

Dr. Y.S.Touloukian added the question "for whom" to Dr. Leake's question "for what." It is essential to distinguish between the generators and users of information, who are distinct in their needs and characteristics.

Dr. Leake agreed that Dr. Touloukian was wise to raise the question, "For whom?" In spite of all that is said about the importance of making contact with those in other fields, scientists are still talking largely to each other. Those who interpret scientific developments for the benefit of people generally may be performing as important a scientific function as scientists themselves. Dr. Leake went on to suggest that an attempt be made to apply measurement techniques to motivations.

Dr. Leonard Carmichael commented that Professor J.D.Bernal in his paper had presented some of the results of his long and thoughtful studies of almost the whole field covered by this Conference. It is written from the point of view of the user and emphasizes the importance of a natural historical approach to the problem. He makes some very concrete proposals as to how the processing of scientific information may be carried out in an effective manner. He points out that knowledge of the requirements of users and of the uses to which they wish to put the information should be the ultimate determining factor in designing methods of storage and retrieval.

Dr. Carmichael went on to say that the point of view of the physiological psychologist, or human engineer, should not be forgotten in considering the problem of scientific communication. For example, the rate of silent reading of standard prose by university students of equally good scholarship has been found to vary greatly (from 2.5 to 9.8 words per second). There are many individual differences in all aspects of information reception and use by human beings, and also many differences in learning and retention abilities. Neurones as well as electronic devices are concerned in the input, storage, search, scanning, synthesis, recall or retrieval, and recognition of scientific information. Human perception and motivation are both better understood today than they were ten years ago, and this knowledge is relevant to the whole problem of the nature and use of scientific information.

A book entitled *Reading and Visual Fatigue,* written some years ago by Dr. Carmichael and Dr. Walter F.Dearborn, was based on prolonged experimentation. Among the findings of this study was the fact that reading of microfilm does not produce more eye fatigue than book reading; nor is the speed of reading of microfilm necessarily different. This finding among others has relevance to the problems discussed by this Conference. Consideration of sensory processes, central nervous system activities, and human motor processes should not be neglected for they are basic in information production and use.

Professor Bernal emphasized that it is important to make a clear distinction between the people who make inquiries and the kinds of inquiries they make. Information about both is basic to the design of information systems. History suggests that the different types of inquirers and inquiries should be analyzed by subject. The habits of different kinds of scientists are dictated by the nature of their work. Separate studies of the part scientific communication plays in every scientific discipline are needed.

Professor Bernal then said there appears to be a very large disparity between the edifice of scientific communication and the practices of research scientists. He had thought that the edifice was built primarily for practical scientists, but the studies indicate that they do not read very much anyway. Possibly an enormous edifice is being built for no particular use at all. Scientific information is pouring out at a rate that is hyperexponential, and reading time, however well it is spent, cannot be increased by a factor of more than two. Something quite drastic has to be done about this situation. There is just too much relevant information. Scientists can perhaps continue what they have always done—work with the aid of hunches, luck, conversations, etc. Something a little more positive, however, ought to be introduced into our scientific information services. More effort might be made to direct scientific information to people who can make use of it. One way would be to make more use of the intelligent information officer who knows the people he serves, as long as he doesn't overdo it and direct their attention to so many things that they are worse off than they were before. Another method would be to go right back to the seventeenth century and use personal contacts more extensively.

Dr. W.F.Libby commented that the Spirit-Kofnovec paper describes a routing system for information by information officers that is fairly standard in many organizations, including the U.S. Atomic Energy Commission, which distributes its reports by categories.

Dr. Libby went on to say that the analysis of reference questions by Herner and Herner contains several very interesting implications. Any arrangement of knowledge must cover to some degree all of the disciplines. The main call on information services, as Professor Bernal has put it, is not in respect of pieces of information that are central to the work of the inquirer, but rather ancillary to it. The worker has means of finding information in his special field of study. The Herners were wise in cautioning readers of their paper that reference questions do not encompass all the requirements of research workers for information. Their analysis of the number of concepts involved in reference questions and the logical relationship between them suggests that the development of information retrieval systems and machines may have been delayed because the designers are trying to anticipate every possible complexity that might be encountered. It may be that relatively simple systems and devices will suffice.

Studying the information needs of scientists is very confusing. Dr. Libby emphasized that such needs vary with the subject, the type of research (pure or applied), the type of user (research worker, historian or writer), the purpose for which the information is sought, the availability of information services, the amount of material on hand, language capabilities and general ability. All these factors make for a confusing situation in studying the information needs of scientists. It appears that all forms of information services are used.

It is apparent from most of the studies that research workers rely greatly on oral communication. Students are required to learn how to use the literature, but it is apparent that as a scientist becomes older and better established he relies more and more on oral communication. It is well known, however, that some of the most productive years of scientists are the younger years. Reliance on oral communication may not be the best technique. On the other hand, the researcher is often seeking information at the forefront of a developing field that is not in the library, and he must get it by letter or conversation or at meetings.

Until now, Dr. Libby said, the abstract journal has not been an ideal medium for searching the current literature because indexes could not be prepared on a current basis. Therefore beginning in January of next year *Nuclear Science Abstracts,* which is published by the U.S. Atomic Energy Commission, will contain author, subject, and report number indexes in each issue. Prompt publication of these issue indexes and of quarterly cumulations of them has been made possible by recent mechanical developments that aid in compiling and publishing indexes. The bibliographic tools, reviews and report services provided by the Atomic Energy Commission demonstrate that the agency recognizes its obligation to make information available in a way in which it can be economically and efficiently used. Some consideration should be given at this Conference to the support and assistance needed in order to make information available.

Mr. Julius Frome concurred with Dr. Libby's remarks on the tendency to make systems overly complicated. In the U.S. Patent Office persons have worried because a certain mechanized system is capable of handling 12 to 50 interrelated factors, whereas the average person using the system has wanted only three or four. Perhaps scientists using such systems should be trained to make better use of them, or perhaps the systems should not be complicated. It is evident in many instances that the average scientist does not want to use a retrieval system as complicated as might be developed.

Dr. Ralph R.Shaw stated that it is not impossible for abstracting journals to publish indexes promptly. It is only a matter of the amount of effort and money we put into such work. This problem is probably simpler than we make it. Scientific communication support has not kept pace with the support of research. Perhaps all we need to do to solve our difficulties is to correct this deficiency.

Sir Alfred Egerton began his comments with a few general remarks. Although there were doubts at the time of the Royal Society Conference in 1948, very few persons now regret that the machinery of publication and dissemination of scientific information was inspected at that time. We can perhaps regret that more was not done to follow up that inspection with vigor and insistence on the needed improvements. It has been wise to place first in this Conference the needs of scientists and technologists. Of the total information extant, only part is in the literature. Much of it is stored in the many brains of scientists and technologists. We are as dependent on biological storage as on mechanical or library storage. The mental processes of abstracting and distillation of scientific literature are of great importance to the advancement of science. The greater the mind that carries out those processes, the more chance there is of effective results, though it is true that certain types of minds have a special facility for such work. The needs of scientists working as a team are different from those of the individual scientist. The latter needs special services that provide up-to-date information in his field of interest and needs copies quickly for his own use of papers he wants to study in detail.

Sir Alfred remarked that the study of the use of scientific literature and reference sources in Denmark and Finland by Miss Tornudd contains a great many detailed findings which indicate a need for the following: easily filed abstracts appearing with the original papers in periodicals; formal instruction in library use on the undergraduate level; international abstract services in specialized fields; reduction of the number of different abstract journals; more good critical reviews; and stricter editorial policies to prevent the rehashing of old material. It was suggested that a study of use of the literature in similar institutions with different library services might be made. Miss Tornudd's study is of interest because it concerns needs in smaller countries where communication by means of the literature is more frequent.

Sir Alfred went on to consider two other papers briefly. The results of a survey of the use of technical literature by industrial technologists, reported by Mr. Scott, indicated that literature is read mainly for general interest and for keeping up to date, rather than for information needed in research. The technologist must be given what he needs rather than what he asks for. Consequently the editor is very important and a key person in the whole information picture. Mr. Herner, in an interview survey of medical scientists concerning use of the literature, found among other things that personal advice from colleagues and face-to-face communication are the most outstanding sources of information and ideas. All of these papers show the importance of the scientist and technologist as a human being with a brain that can be assisted and fertilized by other brains, as well as by the accumulated stock of knowledge in the literature.

Dr. Shaw expressed the view that most of what we hear at meetings of this sort, his own contributions not excepted, is reminiscent of the lawyer addressing the jury who said, “These are the conclusions on which I base my facts.” The Hogg-Smith paper does not do this, but it does not come to any conclusions. His question then is, “So what?” What good do these studies do if they don't tell us what to do after we finish them? Most papers on this subject are made “scientific” by pseudo-scientific statistical windowdressing based on inadequate data. Moreover, scientists are frequently asked to give opinions in areas they really have not studied. It is even worse to ask them to grade their opinions. Dr. Shaw suggested the following questions for discussion:

1. Are there any conclusions to be drawn from these papers that will allow us to take action?
2. Do we not talk too much about operational efficiency and statistics when we should be considering what can be done in a qualitative way to help geniuses and creative workers?
3. Are we assuming that because a man is a great scientist he should be an authority on scientific information services?

Perhaps a contribution of this meeting would be to declare a moratorium on publication of inconclusive summaries of incompetent opinions based on inadequate samples.

Dr. I.H.Hogg agreed with everything Dr. Shaw had said. He went on to say that he and Dr. Smith had undertaken their study because it appeared that the scientists in the Industrial Group of the United Kingdom Atomic Energy Authority were not in fact using the services provided for them. The study did serve a purpose in that it persuaded the authorities that the libraries were much too small and that more Russian translations were needed. Conclusions from the study were not available in time to be included in the conference paper. The conclusions and recommendations from the study have since been prepared and they cover 11 pages.

Mr. Saul Herner commented that although there are many things wrong with the studies that have been done in this area, as there are with all studies, they have contributed something. A start must be made somewhere. An imperfect study is a contribution for it gives a point of departure and a basis for comparison. Everyone has said that we don't know enough from the studies that have been done. No intelligent critical review of these studies has been made. They show a tremendous amount of consistency in some aspects, such as the temporal span of the use of literature. Mr. Herner suggested that we act on the things we do know and put them to test. We will then find out more about what we do not know.

Dr. Herman H.Fussler commented first on Mr. Fishenden's report of a study at the Atomic Energy Installation at Harwell. With respect to the use of volunteer subjects, social scientists might express some concern over the possibility of bias. The paper does not indicate the basis on which the validity of the sample was determined. Broadly speaking, the study seems to confirm others that have recognized the importance attached by scientists to current periodicals as a source of current information, and indicates the scattered and semi-fortuitous use of other techniques and services.

Dr. Stephen H.Spurr in his paper presented a personal account of some of the characteristics of research literature in the field of forestry, and recommended that forestry institutions receive a classified subject file of abstracts and citations in card form; also that specialists in the field of forestry participate in the abstracting and classification of papers. Dr. Fussler remarked that such a technique is clearly feasible if the essential conditions to such a bibliographical approach can be met, namely that a central agency compile the abstracts, or take them from the other sources without impairing the other sources economically, and that there be sufficient funds to sustain this rather expensive form of publication. Card techniques seem most likely to work where there are a relatively small literature and a relatively small group of investigators.

Dr. Fussler turned next to Dr. Urquhart's paper, which presents data relating to the loans of serial publications in 1956 from the Science Museum Library. The data are presented in a terse and abridged form that makes detailed interpretation of some of the findings difficult. The usual concentration of use of scientific serials is strongly confirmed: 80 per cent of the total number of loans is apparently concentrated in less than 10 per cent of the serial titles; of the 9,120 current serial publications received, 4,821 were not borrowed at all; and the most widely held serials are also those most frequently borrowed. Frequency of use, of course, is not the same as a qualitative measure of importance. Dr. Urquhart indicates that in his judgment such data will be useful in shaping; any systematic plan for greater accessibility to scientific serials in the United Kingdom.

The papers in Area 1 had suggested to Dr. Fussler some broad generalizations or observations as follows:

1 It is evident that the communication of scientific information is a complex and variable matter. It is not evident that our knowledge of these variations is yet as complete as it may need to be.
2 It is evident that different techniques and channels are used for different needs and that some are more effective than others.
3 Scientists are likely to learn of major basic developments in their own fields of specialization rather quickly and easily. There is less assurance concerning peripheral information within a specialized field and still less concerning relevant information from other fields.
4 There is rather substantial evidence that scientists are not in many cases notably systematic in covering the literature, and indeed the existing tools are not exploited by users to anything like their full potential.
5 Chance associations of ideas and information appear to be important. If this is true, we should be giving thought to information handling devices that would increase the probability of such events.
6 The studies thus far do not advance an entirely satisfactory explanation of the apparent unsystematic use of available sources and services.
7 The study of the behavior of scientists in relation to their literature need not be pressed to the last unknown. To do so would require research techniques infinitely more sophisticated than those employed thus far. It should be possible soon, if not now, to begin to outline the specifications for a series of feasible information systems. The specifications should take into consideration three major elements: the habits and needs of scientists as broadly conceived, the economic and material resources that affect the distribution of information, and the characteristics of scientific literature itself.

Mr. R.M.Fishenden said that he believed the results of his survey at Harwell give a fair picture of what is going on and of the failures in the communication system. One general point made by nearly all the papers is that hardly any research worker has access to all the potentially useful information that an information worker may think he ought to have. What we don't know is whether it matters. It seems that we should direct attention to this question next. What is the penalty, for example, for not having access to the Russian literature or for not bothering to read through all the abstract journals and lists available?

Dr. D.J.Urquhart explained that statistics on the loans of scientific periodicals were compiled for a very practical reason: a library is being built and there was need to know how many copies of publications to acquire. He would like to receive the results of any other use studies that would help in deciding what publications to acquire for the new National Lending Library of Science and Technology in England.

Dr. Urquhart indicated that this business of measuring what scientists do should not be overdone. The Department of Scientific and Industrial Research in London has made a number of studies and has published some results, but it appears now that the least important results have been published. Whenever a measurement determined policy, it was acted upon and there was no need to publish it. For many purposes crude, simple measurements can be quite useful. We do not necessarily want highly developed experimental techniques. Even crude questionnaires have produced results that have been accepted and acted upon.

The problem of language is a subject Dr. Urquhart thought the conference might seriously consider. It is fortunate that a great deal of the scientific information happens to be in the English language, but this is an accident of history. There is nothing permanent about it. We should seriously consider what language, real or artificial, we can agree to have as a secondary means of communication throughout the world, a language we should all learn.

Dr. Eric de Grolier stated that the results of a questionnaire survey of 200 scientists in France support very strongly the belief that we have a situation of over-communication and that scientists see a need for stricter editorial policies in order to eliminate papers that are not useful. The results also suggest that scientists would favor a drastic reduction in the number of different abstracting services. A clear tendency toward the development of more personal contacts among scientists is discernible. The full report of the study will be published by the French Union of Documentation.

Mr. Herbert Ohlman suggested that since it is difficult for people to remember what they do and use, perhaps we need a system of tallies on the actual pages of materials used, which could be summarized at the time a library makes its inventory. We might even encourage marginal comments in the publications.

Mr. Eugene Garfield expressed the view that too many people are afraid to have opinions. They have to conduct extensive surveys before they are willing to express opinions. If John Shaw Billings had conducted a survey before he went ahead with the Index Catalogue of the Library of the Surgeon General's Office, we might never have seen it. The phrase “information requirements” is deceiving. It is confused with the information-gathering habits of scientists. The papers in this area do not cover information requirements as such. Mr.

Garfield's firm recently did a very small inquiry of about five organic chemists and learned that they need and want specific information about the new organic chemicals that are being made.

Dr. Leake commented that scientific productivity is a creative process, which depends upon many factors opposed to stereotyping. Scientists should devise some way of keeping abreast of important developments that occur in fields other than their own and should make some provision for inspiration from either the unusual or the unknown.

Dr. Morse, in bringing the discussion to a close, observed that it is a little surprising that there has not been a review or study of the role of the scientific meeting in the distribution of information. Perhaps such a study should be made. The conclusions in the studies made to date have usually been limited to specific problems. General conclusions of wider interest are rather more difficult and will result from further studies that will be and should be made.

HELEN L.BROWNSON, *Rapporteur*
PHILIP M.MORSE, *Discussion Panel Chairman*

AREA 2

THE FUNCTION AND EFFECTIVENESS OF ABSTRACTING AND INDEXING SERVICES

AREA ORGANIZATION

Authors of Papers

Members of Discussion Panel

Chairman: ELMER HUTCHISSON, American Institute of Physics, New York, N.Y.

J.D.BERNAL, Department of Physics, Birkbeck College, University of London, London, England

G.A.BOUTRY, Abstracting Board, International Council of Scientific Unions, Paris, France

G.MILES CONRAD, Biological Abstracts, University of Pennsylvania, Philadelphia, Pa.

E.J.CRANE, The Chemical Abstracts Service, The Ohio State University, Columbus, Ohio

S.H.GOULD, Mathematical Reviews, American Mathematical Society, Providence, R.I.

WILLIAMINA A.HIMWICCH, Galesburg State Research Hospital, Galesburg, Ill.

GEORGE SHORTLEY, Booz-Allen Applied Research, Inc., Bethesda, Md.

ALLEN J.SPROW, Psychological Abstracts, American Psychological Association, Washington, D.C.

MAURICE B.VISSCHER, Department of Physiology, University of Minnesota, Minneapolis, Minn.

PROPOSED SCOPE OF AREA 2

FOR A SCIENCE to become and remain completely healthy, information which one scientist discovers or develops must be available under "controlled conditions" to any other scientist who wishes to obtain it—whether to apply in his own research, to use in teaching, or simply to keep informed of progress in the field. ("Controlled conditions" as used here means merely a system of processing and storage that permits ready identification of and access to desired information. It has no implications of granting or refusing permission to obtain information.) The principal system that has been evolved over the decades for the accomplishment of this goal has as its chief components the journals of original publication, abstracting and indexing services, and reviews —in short, the media of primary and secondary publications.[1]

From one point of view primary publication is an early stage in the process of storing scientific information. However, the primary reason for publishing scientific journals, laboratory reports, and other first accounts of research work is to disseminate scientific information—not to store it. The steady and relatively unorganized growth of primary publications in recent years has greatly increased the scope and complexity of the storage and retrieval problem. With reference to secondary publications, abstracts, indexes, bibliographies, and so forth, some scientists look upon these primarily as guides to current publication; for other scientists, however, they are primarily storage and retrieval devices. This Conference will devote most of its attention to the latter aspects of secondary publications although some consideration will undoubtedly be given to the function of disseminating information because all of these functions are so closely related.

Of course there are many kinds of abstracts and indexes. Some are comprehensive, such as *Chemical Abstracts;* some are specialized, such as *Petroleum Abstracts.* There is a great deal of difference in services and in coverage. For example, Gray and Bray found over 150 abstracting services that appeared to be of interest to physicists. On the other hand, many areas are not served at all, at least in the English language.

The existence of satisfactory abstracting and indexing services seems to depend in the first instance on the money available. Thus the fields of chemistry

[1] The proposed scope of the Conference Area, as shown here, was prepared during the Spring and Summer of 1956 and provided to all potential contributors as a guide to the aims of the Conference.

and medicine are very well served. Accidental factors in historical development may also play an important part. Chemists are served by an abstract journal, whereas engineers rely primarily on index cards. It would be interesting to examine the question to ascertain whether this difference really grows out of a difference in needs, or whether it is primarily a historical accident.

There are a number of approaches to the problem of supplying complete abstracting and indexing services for a field. One is to establish a centralized service like the All-Union Institute of Scientific and Technical Information of the Academy of Sciences of the USSR. A second is the approach adopted by the ICSU Abstracting Board. A third is to establish specialized services employing professional abstractors. The Russian Institute states that it furnishes complete coverage of science on the basis of 10,000 journals. From these 10,000 journals they prepare 400,000 abstracts, which is to be compared with 80,000 that *Chemical Abstracts* prepares on the basis of covering 7,000 journals. The Russians agree with the *Chemical Abstracts'* decision that professional abstractors should not be used, and employ 13,000 working scientists to prepare the abstracts. The abstracts may appear in any or all of the 13 abstract journals issued by the Institute. Abstracts may be prepared specially for the appropriate abstract journal where this appears necessary, or they may be duplicated in the different journals.

The ICSU cooperative program encourages the use of author abstracts. Every effort is made to achieve promptness by making special arrangements with scientific journals to furnish advance proofs of articles. There are many examples of small or specialized services using professional abstractors.

It appears that it would be very interesting to compare the advantages and disadvantages of all these approaches from the standpoint of cost of the service, promptness, quality of abstracting, and so forth.

PRINCIPAL TOPICS FOR DISCUSSION

Research into a number of aspects of abstracting and indexing services is needed. Among typical problems that may be examined at the Conference are the following:

1 What measures of effectiveness can be devised to determine the adequacy of abstracting and indexing service as storage and retrieval devices? It might be noted here that a study by the Department of Scientific and Industrial Research, based on their collection of questions that could not be answered readily by a search of the literature, has indicated that although a very large proportion of the important papers are covered by the abstracting services, many of them cannot be located because of inadequate subject indexing. This whole problem of providing adequate

subject approaches to abstracting journals, or any other medium for storing information, should be given special study.

2 What measures of effectiveness can be devised to evaluate the differences between indexes, lists of annotated titles, various types of abstracts (descriptive, indicative, and critical), and reviews, in terms of the information carried and the effort involved in their preparation and use?

3 What is the optimum amount of information that should be contained in abstracts and indexes to render them effective as storage and retrieval devices, and upon what variables does this optimum depend? For example, ease of use, completeness, cost of preparation, speed of appearance, type of scientific field, and so forth.

4 What useful new services might be developed based upon the existing resources presently available at abstracting and indexing services? These might include film editions coded for use with rapid scanning devices, coded tapes, special reference service for requestors, and so forth. Also, how can the cost of these services relative to their value to requestors be determined?

5 What are the advantages and disadvantages of a centralized organization such as the All-Union Institute of Scientific and Technical Information of the Academy of Sciences of the USSR? To what extent can duplication in scanning be eliminated? To what extent can multiple use of abstracts be made among disciplines, and to what extent must duplicate abstracts be prepared? What cost data is available on scanning, preparation of abstracts, editing, and so forth, under such a system?

6 In comparison with No. 5, what are the advantages and disadvantages in terms of the scientific effort involved in a cooperative program using author abstracts along the lines of the ICSU Abstracting Board experiments? A study of the ICSU Abstracting Board's cooperative program for preparing abstracts in the field of physics should throw some light on this.

7 What is the effect on storage and retrieval systems, especially mechanized systems but also manual systems, of variations in vocabulary among abstractors and from field to field in interdisciplinary areas such as instrumentation?

8 To what extent are abstracts now regarded as means of dealing with the foreign language problem?

9 Is there any feasible method by which the abstracts and indexes prepared in one field of science might identify information (e.g. instrumentation or mathematical techniques) useful in another field?

An Evaluation of Abstracting Journals and Indexes

MAURICE H.SMITH

The purpose of this investigation is to examine the relation of typical abstract journals and indexes in the physical sciences and engineering to some literature search activities in a basic research situation. The structure of the literature and the characteristics of abstract journals and indexes in relation to its structure are discussed.

Since the work described was carried out at the request of engineers and scientists, and was accomplished with their close cooperation, it is hoped that their needs will be reflected to some extent in the results. The remarks of engineers and scientists on the problems of literature searching and of keeping abreast of their fields and locating information are included in the discussion informally.

All the work described has been done at the James Forrestal Research Center of Princeton University. The principal research programs involved are in the fields of jet propulsion, low-speed aerodynamics, gas dynamics, and flight dynamics of helicopters and other airborne vehicles. Some literature search work has been concerned with aspects of research programs in chemical kinetics, metallurgy, and engineering phases of nuclear physics programs.

The location in 1951 of the University's Department of Aeronautical Engineering and its laboratories at the Forrestal Center, five miles from the campus, caused the establishment there of its departmental library, which was merged with the collection of reports known as the Project Squid Library. Project Squid, a cooperative project in jet propulsion, has its headquarters at the Forrestal Research Center. The two merged collections were augmented by interested members of the faculty who gave sets of chemistry and physics journals, and became the Forrestal Research Center Library (9). It is one of the special collections of the Princeton University Library.

The presence of the research programs in the scientific fields required that the more commonly used journals and books in chemistry, physics, and applied

MAURICE H.SMITH The James Forrestal Research Center, Princeton University, Princeton, New Jersey.

mathematics should be available. The book and periodical collections at the end of 1957 included approximately 60 per cent basic science material and about 40 per cent aeronautical engineering material. Eighty journals are received, and the collection includes about 2000 books and 1600 volumes of periodicals.

The collections of reports include about 30,000 separate security unclassified titles, and approximately the same number of reports on microcards received from the Armed Services Technical Information Agency. Aeronautical engineering is strongly represented, with fewer reports in the basic sciences, including about 1000 atomic energy reports.

The card catalogs cover books and reports by subjects and authors. Separate papers in the proceedings of congresses and similar collected volumes are indexed.

Because of the existence elsewhere in the University Library system of departmental libraries, the collection at Forrestal must be selective. The selection of materials for purchase is made by the faculty, and is based also on research demands. There is a Library Committee which advises on major acquisition and operating problems. The Committee at present includes two aeronautical engineering professors, the chief engineer of the jet propulsion projects, a recent doctoral graduate associated with jet propulsion research, a metallurgist, two nuclear physicists, and an engineer associated with nuclear research. All major research areas are represented.

LITERATURE SEARCHES AND DATA PRESENTED

Literature search requests are made on behalf of research projects. The report on results may be a stack of assembled material, a typed list, or an offset printed bibliography with abstracts, and author, subject, and source indexes.

The usual procedure is to assemble the references on 5 by 8 inch cards which accommodate abstracts or notes. In recent cases the means by which the reference was located is recorded on this card in a suitable brief manner. In the Appendix will be found the cases selected for use as basic data.

It is felt that the diversity of requests and of ways to satisfy them is reflected to some extent in the 50 cases presented in this paper. There is one previous paper in the literature (3) containing a smaller sample of bibliographies compiled, which approaches the problems of literature search in a somewhat similar manner.

The fifty cases will be seen to group themselves into three general types, which may be designated as comprehensive, intermediate, and brief. Their grouping begins with compilations requiring weeks of time, on behalf of more than one person, usually including abstracts, usually printed by offset, and usually equipped with indexes of authors, subjects, and possibly report sources. The number of references may be as few as about 70, or as many as 350, which is approximately the upper limit for a modest service to undertake. The group of comprehensive cases (1 through 10) answers to this general description.

The intermediate group (cases 11 through 37), are less complete, were usually accomplished in days rather than weeks, usually have no abstracts, may have been done for either a group or an individual, usually are typewritten, but may be duplicated by Ditto or other informal method, or may be recorded only on cards or in manuscript note form. The great variation within the intermediate group is further indicated by the number of references, ranging from a half dozen to more than 100. Two cases, 11 and 28, dealing respectively with *Taylor Instability* and with *Tables of Kelvin Functions,* were felt to be reasonably complete at the time of compilation. For the most part, however, the intermediate group represents only the best results obtainable within a limited amount of time.

The final group of cases (38 through 50) consists of some instances of brief searches for specific information, solved fairly rapidly. The number of references in these is small, ordinarily not exceeding 25, and more often less than 10. The records are never more formal than a typed memorandum, and frequently are much less formal than that.

It is perhaps appropriate to mention that the mechanics of recording literature search work is similar to that of making research notes, and is equally important in the conduct of the work. Entire searches can evaporate without methodical recording. In the sample used in the present study, Cases 23, 34, 40, and 46 were never formally summarized, except in periodic reports on literature search work. Details of cases 23 and 34 were reconstructed from memory for the purposes of the present study.

TYPES OF LITERATURE AND THEIR USE

The greater part of the literature represented in the sample of cases consists of articles in journals publishing original research, and technical research reports. A third category is articles in collected volumes of papers, such as the proceedings of congresses and symposia, commemorative volumes, survey volumes, and annual review volumes.

In 15 cases readily assessed for the types of literature represented in them, there were recorded 601 journal articles, 1048 reports, 208 articles from collected volumes, 17 books and 43 of such types of literature as engineering society preprints, dissertations, and patents.

These figures are indicative only as a representation of the use of research materials by research personnel. They reflect the fact that the literature recorded in these particular examples consists of these proportions of types of literature. The point made is that journals alone obviously do not carry the research load, that they should not carry it, and that the entire spectrum of the recorded available literature must be considered in relation to information problems.

For a more realistic view of the actual use of the principal types of scientific and technical literature, reference is made to Table 1, "Unclassified Items on Loan, November 22, 1957." Attention is called to the limitations of this sample of use. It does not distinguish clearly between the use of types of literature for teaching and study, compared with research. It is impossible to do so because of the participation of graduate students and professors in research programs, and the participation of research associates in curricular pursuits, both in classroom courses and in laboratory study. Nor can it show the use of atomic energy report material by nuclear energy projects, received directly by them, the use of the University's other departmental libraries in chemistry, engineering, and physics, nor the use of the general collections of the University Library.

TABLE 1 The Forrestal Research Center Library of the Princeton University Library: Unclassified Items on Loan, November 22, 1957

	Books		*Periodicals*		*Reports*		*Total*	
Aeronautical engineering								
Graduate students	104		13		86		203	
Faculty	71	17	174	262				
Subsonic	22	11	288	321				
Helicopter and instrumentation	79	31	117	227				
Gas dynamics	75	29	103	207				
Jet propulsion	75	13	195	283				
Total		426		114		963		1503
Matterhorn	185		32		31		248	
Metallurgy	22	10	9	41				
Organic chemistry	33	7	—	40				
Mathematics	7	—	4	11				
Accelerator	36	—	—	36				
Total		283		49		44		376
Campus students	18		3		31		52	
Campus faculty	3	10	27	40				
Other campus	3	—	5	8				
Total		24		13		63		100
Outside interlibrary loans	—			9		76		85
Totals		733		185		1146		2064

It does reflect a pattern of use which has been fairly constant since 1951: that is, a heavy use of books compared with the use of journals by research personnel, a very high use of reports by research personnel, and a high density of items per person in aeronautical engineering (an average of about 18 items per person).

In relation to abstract journals, a distinction between published and unpublished reports is perhaps useful. The report literature is new in recent years only in relation to its quantity (4), for it has been in existence for some time. In the aeronautical field, it has been the chosen form of publication by the national research laboratories from the beginning. Some examples of numbered series of long standing are:

France. Ministère de l'Air: *Publications Scientifiques et Techniques.* 1930 to date.

Great Britain. Aeronautical Research Council: *Reports and Memoranda.* 1909 to date.

U.S. National Advisory Committee for Aeronautics: *Reports.* 1915 to date.

These are priced, published reports. Many series are not priced, but are widely enough distributed and readily enough available from the source that they are considered to be published reports. For example, the *Technical Notes* of the U.S. National Advisory Committee for Aeronautics have long been indexed in the *Engineering Index* and the *Aeronautical Engineering Review;* and the *Information Circulars* and the *Reports of Investigations* of the U.S. Bureau of Mines have been included in *Chemical Abstracts,* as are at present many reports of the U.S. Atomic Energy Commission. The distinction between published reports and unpublished reports is one of reasonable availability and reasonable price. Many atomic energy reports are available from the U.S. Office of Technical Services at reasonable prices and therefore can be considered published material. The same is true of some reports of agencies of the U.S. Department of Defense. But considerable quantities of reports are available generally only through the U.S. Office of Technical Services at microfilm and photostat prices and, for this reason, are considered to be unpublished material. They are of course available free to contractors of the Government who require them, but the body of less-known material in report form is formidable (1, 4) and is capable of being missed to some extent because of the complication of its bibliographic control.

It will bear repeating in relation to types of literature that the entire spectrum of primary source materials, including journals, reports both published and unpublished, engineering society preprints, dissertations, masters' theses, patents, manufacturers' bulletins and translations, must be looked for at one time or another in a basic research situation; and many of them must be stocked as a matter of routine. Secondary sources, including books and trade journals, play a more important part than might be supposed.

Special mention should be made of an increasing type of primary and secondary source material which often requires systematic direct examination. Collected volumes of papers, whether an annual review volume, the proceedings of a congress, or a commemorative volume, are on the increase in both the book and the report literature. The papers vary considerably in quality, but frequently the author's contribution is available only in this form. The abstract journals, understandably, are likely to cover them late. The *Engineering Index* and the *Aeronautical Engineering Review* list their contents but do not abstract the papers. In the work described in this paper, access to this type of literature has been largely through the library catalog and by direct examination.

Some examples of this type of book are the review volumes, such as *Advances in Applied Mechanics, Advances in Catalysis, Annual Review of Nuclear Science, Annual Review of Physical Chemistry,* and *Reports on Progress in Physics;* the proceedings of such congresses as the International Congress of Applied Mechanics, Midwestern Conference on Fluid Mechanics, Heat Transfer and Fluid Mechanics Institute, National Electronics Conference, Japan Congress of Applied Mechanics, U.S. National Congress of Applied Mechanics, Symposium (International) on Combustion, Anglo-American Aeronautical Conference, or the various conferences of the Advisory Group for Aeronautical Research and Development (AGARD) of the North Atlantic Treaty Organization.

Some more esoteric examples are:

Actes du Colloque International de Mecanique, Poitiers, 1950. Published in *Publications Scientifiques et Techniques du Ministère de l' Air,* Paris, Nos. 248, 250, and 251 (1951).

Memoirs sur la Mecanique des Fluides offerts à M.Dmitri P.Riabouchinsky. Published in *Publications Scientifiques et Techniques du Ministère de l' Air,* Paris, Hors Serie, 1954.

Boundary Layer Effects in Aerodynamics. Proceedings of a symposium held at the National Physical Laboratory, March 31 and April 1, 1955. London, H.M. Stationery Office, 1955.

RELATION OF ABSTRACT JOURNALS AND INDEXES TO LITERATURE SEARCHING

The results of 41 of the 50 cases would have been impossible without the use of abstract journals and indexes. In 6 of 10 comprehensive bibliographies, containing 1047 references, 537, or 51 per cent, were identified through abstract journals and indexes. In 12 of 27 intermediate cases, containing 188 references, 130, or 69 per cent, were located through abstract journals and indexes; and 100 per cent of the references in 9 of 13 brief searches were located by this means.

Judgment must be used on which abstract journals or other bibliographies are likely to be productive, or whether some other approach is more appropriate for a particular case. It appears from the evidence that other techniques must be combined with the judicious use of abstract journals and indexes for the greatest possible efficiency.

For certain types of information, it is more efficient to use other sources and techniques entirely, such as asking authoritative persons, direct examination of likely items and, above all, the location of one or more review articles, one or more recent articles having bibliographies, an existing bibliography bearing on the subject, a treatise, a monograph, or a dissertation. In cases 40 and 41, for example, somebody was asked. In cases 14, 15, 18, 19, 23, and 26, the literature was consulted directly; and in case 16 there was some use of bibliographies after the greater part of the literature had been recorded by direct examination of the principal available journals and congresses in the field of the inquiry, astronautics.

It will be noted that several of the 50 cases fall in fields which are of interest to more than one discipline. For example, case 1, *Origin of Turbulence, and Related Subjects,* has engaged the interest of physicists, mathematicians, chemists, chemical engineers, mechanical engineers, aeronautical engineers, and meteorologists. Case 6, on the atomic rocket and nuclear propulsion, has interested nuclear physicists, chemical engineers, and aeronautical engineers. Case 7, on liquid ozone-oxygen mixtures as rocket oxidizers, has interested chemists and aeronautical engineers. Case 11, *Taylor Instability,* has interested physicists, nuclear physicists, chemists, and aeronautical engineers. Such interests by both scientific and engineering disciplines naturally are reflected in the literature and will require the use of abstract journals and indexes accordingly. This accounts in part for the use of two or more sources in 9 of the 10 comprehensive searches, 13 of the 27 intermediate searches, and in one of the brief searches.

Another more serious reason for using more than one source is that the information may lie in a field which is too new for its literature to be under complete control, or the type of literature involved is not widely available and is not under complete control. The publishing lag for abstracts is a contributing factor. Case 4 perhaps illustrates these points in the field of helicopters, or case 5, on vertical take-off and landing aircraft. A number of cases illustrate the importance of the report literature which has been previously discussed.

It would seem that duplications will always exist among abstract journals as long as they index the same material on behalf of disciplines which have overlapping interests. Duplications on thermodynamic properties (case 7), for example, may be expected between *Chemical Abstracts, Physics Abstracts,* and various engineering sources. Wide duplication may be expected among scientific and engineering abstract journals and indexes in such fields as fluid mechanics (cases 1, 8, 11), and combustion (cases 7, 17, and 22).

SOME COMMENTS OF ENGINEERS

Discussion of the results of some of the completed work with the persons who requested it brought out some interesting points. There was a general disinterest in discussing finished business, especially in comprehensive cases which had already taken up considerable of the engineer's time during their compilation.

There was considerable interest in the date when certain of the searches (cases 3, 7, and 8) would be finished, and whether there was anything new since progress on them had last been discussed.

An electronics engineer who had received two articles in bound volumes (case 49) had found some useful information in the one which arrived promptly. The other arrived a day later. He could not see why it would be on his desk and sent it back.

Another electronics engineer pointed out that he had combined and greatly enlarged the bibliographies of cases 25 and 26, Correlation Analyzers and Power Spectrum Analysis. Case 25 had not been done at his request, but he had acquired a copy. This engineer made the valuable suggestion in relation to case 50 that the searcher should insist upon a full definition of the problem, and should not begin unless there is sufficient information. The references first submitted in case 50 were useless because the problem had been too broadly stated. The British report mentioned in the description of the case had evidently been located through intuition, as there was no way of knowing from the request that it would be the exact kind of thing needed. Insistence upon a clear understanding of exactly what the inquiry is about, and the purpose for which it is to be used, is of course essential to an efficient approach to the problem (7). In the 50 cases, the titles are those of the inquirers and they were stated very clearly at the time of the inquiry, with suggestions on possible authors, laboratories, and sometimes with several sample references. A more exact definition had to be obtained in case 49, *Magnetic Clutches for Servomechanisms,* and one of the two inquiring chemists in case 9, on the water-gas reaction, amplified and clarified the original question.

It will be clear that the literature searching described is only a part of a body of activity involving some 200 professional scientists and engineers. Most of them maintain bibliographical files of some kind, and personal collections of books, reports, files of professional and trade journals, photostats of articles, reprints, preprints, and correspondence. A senior professor of the caliber to give invited papers at international or national congresses and symposia will of course have a larger collection than a junior colleague of more modest attainments, but all have some kind of personal collection.

One chemist, two electronic engineers, and eight aeronautical engineers were asked informally about their personal methods of keeping up to date and of finding information.

The chemist, six aeronautical engineers, and both electronics engineers felt that the location of one or more reliable recent journal papers or reports was their most frequent method of finding information. Six aeronautical engineers mentioned use of *NACA Research Abstracts* (U.S. National Advisory Committee for Aeronautics), and five used also that organization's *Index of NACA Technical Publications.* Two mentioned *Applied Mechanics Reviews,* one mentioned *Mathematical Reviews,* and one who used both of these mentioned the usefulness of *International Aeronautical Abstracts,* published since January 1956 in the *Aeronautical Engineering Review.* The chemist regularly uses *Chemical Abstracts,* and felt that no other abstract journal could equal it in breadth of coverage, quality of subject and auxiliary indexing, and general excellence. He mentioned its author and title indexes. One aeronautical engineer felt that the titles of papers in the author index would save considerable of his time. One aeronautical engineer who uses the *Bulletin Signaletique du Centre National de la Recherche Scientifique* felt that none of other sources he consults could match its coverage. He was a strong proponent of the value of meetings and conferences, not so much for the material on the program as for the opportunity to talk with colleagues. Three other aeronautical engineers made special mention of the value of contacts with colleagues in this manner.

The two electronics engineers valued *Electrical Engineering Abstracts* highly. One used *NACA Research Abstracts* and certain company library accessions lists, in addition to the Forrestal Research Center Library accessions list. Three aeronautical engineers also mentioned the library accessions list.

All were agreed that there is no procedure applicable to all problems of information. One aeronautical engineer felt that abstract journals were too time-consuming. He maintains an extensive card file, and spoke of the usefulness of the German perforated card service issued previous to World War II by the Zentrale für Wissenschaftliches Berichtswesen über Luftfahrtforschung (ZWB).

While no conclusions are attempted on such a small body of evidence in regard to the information-gathering habits of scientists and engineers, the impression from the body of experience reported in this paper is that senior engineers of great attainments are indistinguishable from scientists of like age and experience in this respect. Both appear to react the same to complex and inadequate bibliographic apparatus. Both are likely to ask for help when confronted with a problem outside their immediate fields involving such apparatus, and both appear to have a full knowledge of the values of the literature from long exposure to it. The most analytical of scientists has declared himself helpless when confronted with a time-consuming bibliographic task on top of a busy schedule, and the most empirical of engineers has demonstrated complete analytical powers when supplied with a few starting points. It is difficult to generalize on the matter (6).

COVERAGE, QUALITY OF ABSTRACTS, AND SUBJECT INDEXES

The three qualities of abstract journals and indexes which are most difficult to evaluate are coverage, quality of abstracts, and subject indexes.

There is no definite proof in the present paper that any specific abstract journal lacks coverage in its field, that its abstracts are lacking in quality, or that its subject indexes are inadequate. The detailed checking and analysis required to make such evaluations are outside the scope of the paper.

The user of abstract journals and indexes must believe that the information is there to be found, provided that a realistic diagnosis of the information problem has been made, and provided that the problem is approached with a reasonable knowledge of what each abstract journal or index claims to cover. Occasionally, the results will be negative, or a single abstract journal will not do what can be done by using two or three.

For example, the inquirer in case 1, *Origin of Turbulence,* was interested in a paper by Hermann Schlichting on the subject. The particular paper desired was “Uber die Theorie der Turbulenzentstehung,” in *Forschung auf dem Gebiete des Ingenieurwesens,* v. 16, No. 3, 1949/1950, pp. 65–78. No index appears to have picked it up, and it was located eventually by going back through the bound volumes of the journal.

In the same case, all the sources consulted together did not turn up all of the 11 papers on the subject by L.Sackmann. One was located by direct examination of one of the three volumes of the *Actes du Colloque International de Mecanique,* published in the numbered series of *Publications Scientifiques et Technique du Ministère de l' Air* in 1951. Certain of the papers published in 1947 and 1948 in *Comptes rendus* were recorded by no abstract journal or index.

In another case, not one of the 50 of the present paper, some work by the British scientist, P.C.Thonemann, was of interest. A list of eight journal articles, six letters to the editor, and one research report was compiled. The procedure was to check *Nuclear Science Abstracts,* 1948 through December 15, 1956, *Physics Abstracts,* 1944 through November 1956, and *Chemical Abstracts,* 1943 through November 15, 1956. These turned up six, five, and four items, respectively. No attempt was made to see how many of the total of 15 were covered by *Physics Abstracts* or *Chemical Abstracts.* The point is that all three must be searched for a reasonable assurance of completeness. Special mention should be made of the general acceptance of *Chemical Abstracts* as the most comprehensive journal of its kind.

If the structure of the literature is considered, *Chemical Abstracts* leads the way with systematic coverage of books, congresses, journals, patents, published reports, published translations, patents, dissertations, and a certain amount of unpublished material. *Nuclear Science Abstracts* includes all these, as well as unpublished reports and translations. It is perhaps less comprehensive on dissertations, as principally those accomplished under Atomic Energy Commission projects are believed to be included. It should be noted that foreign atomic energy reports and reports of nuclear interest from non-Atomic Energy Commission sources are covered by *Nuclear Science Abstracts.*

The *Aeronautical Engineering Review-Aeronautical Engineering Index* includes the entire spectrum of the literature except dissertations and patents. *Applied Mechanics Reviews* is similar in coverage of various types of literature, but is a critical review and is much more selective on unpublished material. The *Engineering Index* does not cover dissertations, patents, and unpublished reports and translations, and is selective toward the practical side in its coverage of published material, as *Applied Mechanics Reviews* is selective toward the analytical side. *Electrical Engineering Abstracts, Physics Abstracts,* and *Mathematical Reviews* would not be expected to cover dissertations, patents, or very much unpublished material.

The quality of abstracts may be partly evaluated by external evidence. Abstracts signed with the full name and country of the abstractor *(Applied Mechanics Reviews, Mathematical Reviews)* or by the full name or initials of the abstractor, with a frequently published list of abstractors *(Chemical Abstracts),* afford a good prognosis. The evaluation of the work abstracted by the specialist abstractor in *Applied Mechanics Reviews* is especially useful. The inclusion of some signed abstracts with a preponderance of authors' abstracts is characteristic of *Nuclear Science Abstracts* and the two parts of *Science Abstracts* (*Physics Abstracts* and *Electrical Engineering Abstracts*). When abstracts are staff-written, as in the case of the *Engineering Index,* they are uniform but are descriptive rather than analytical. The staff-written abstracts of the *Aeronautical Engineering Review* are probably based upon the authors' abstracts. They vary from full treatments to descriptive phrases. The author abstracts in *International Aeronautical Abstracts,* contained in the *Review,* are uniform.

The author abstracts in *NACA Research Abstracts* are sometimes condensed. The author or contractor abstracts in the *Technical Abstract Bulletin* (U.S. Armed Services Technical Information Agency) are very full, as are those in the earlier *Technical Information Pilot* (U.S. Library of Congress. Navy Research Section). The author abstracts in *Dissertation Abstracts* are very full. The author or contractor abstracts of U.S. *Government Research Reports* are uniform and adequate.

Comprehensiveness leads to great size, and when the subject index is of great excellence, its compilation is correspondingly delayed. *Chemical Abstracts* would seem to be now in a dilemma caused by two of its characteristics which set a very high standard (2). A prompt excellent subject index and comprehensive coverage are not compatible as judged by the *Chemical Abstracts* operation.

The subject indexes of *Applied Mechanics Reviews, Nuclear Science Abstracts, Physics Abstracts, Electrical Engineering Abstracts,* and *Mathematical Reviews* are prompt and are invaluable to the user. It is felt that mathematicians do not make heavy use of subject indexes. This was confirmed by one aeronautical engineer of strong mathematical bent. The subject index of *Physics Abstracts* has been called inadequate (8).

The *Engineering Index* is in itself an excellent subject index. It is increasingly late in appearing. In 1942 it was likely to arrive in May of the following year. In 1957 it was likely to arrive in August. The *Aeronautical Engineering Index* is made up from the abstracts included in the *Aeronautical Engineering Review* each month, which are classified by a compromise between broad and detailed subjects. The *Aeronautical Engineering Index for* 1956 has been announced for May 1958. Beginning in January 1956, the *Review* contains *International Aeronautical Abstracts,* an offset printed insert on yellow paper of abstracts of original research contributions, both reports and journal articles. The less selective regular letter-press printed abstracts have appeared as usual. There is duplication between the two sets of abstracts, the letter-press abstracts appearing a month or so later.

The *Technical Abstract Bulletin* contains a subject index in each issue, announced to be cumulated semi-annually. The subject index to *U.S. Government Research Reports* appears every six months. The *Technical Information Pilot* had a subject index in each issue, and cumulated subject indexes without abstracts were issued for the years 1948, 1949, 1950, and 1951, appearing rather late. *NACA Research Abstracts* has no indexes, but the *Index of NACA Technical Publications* appears annually, without abstracts. The *Title Announcement Bulletin,* replaced by the *Technical Abstract Bulletin,* had no subject indexes and no abstracts.

SOME ADDITIONAL ABSTRACT JOURNALS AND INDEXES

For various reasons, certain indexes have been omitted entirely from the discussion of this paper, or have been mentioned only briefly.

Two of these are card services. The weekly card service of the *Engineering Index,* available in subject categories at various prices, was rejected by the Forrestal Research Center Library Committee, on the basis of cost and of interest in a more selective attention to basic research material. For similar reasons, the weekly (or more frequent) card index of the Pacific Aeronautical Library of the Institute of the Aeronautical Sciences was rejected, as was the *Industrial Arts Index,* announced to be issued in 1958 in two parts, science-technology and business.

The *Translation Monthly* (the John Crerar Library and the Special Libraries Association), including material formerly covered by the discontinued *Bibliography of Translations from the Russian Scientific and Technical Literature* (U.S. Library of Congress), is mentioned here because of its importance in the literature spectrum.

There is no mention in the paper of the array of Russian abstract journals, as there is no basis of experience by the writer upon which to discuss them.

Dissertation Abstracts has been mentioned only sparingly. The *Monthly List of Russian Accessions* (U.S. Library of Congress) has been mentioned even less. These are available in other parts of the Princeton University Library and have occasionally been used to supplement the general abstract journals and indexes.

CONCLUSIONS

The importance of abstract journals and indexes in a basic research situation is considered to be proved. Other techniques of searching cannot supplant their use, and there is no substitute for them as matters stand now (5). A specialized information center must use them as a basis for setting up a codified body of knowledge (10). Of eight principal abstract journals and indexes discussed in the paper, six are published by scientific and technical societies. What follows is a suggested standard for abstract journals, offered in the belief that societies have demonstrated an interest in professional standards (11) and are concerned with their maintenance. There is strong evidence that government and private agencies have similar interests.

Coverage, quality of abstracts, and subject indexes have been discussed in the body of the paper. The more mechanical characteristics of abstract journals are added to these in the proposed standard. There is no intention of implying, however, that quality is a mere mechanical matter. It may seem that some characteristics suggested for a standard are mutually exclusive, for example strong coverage and a prompt subject index. Such problems have been at least partially solved, and the fact that they are difficult seems to be no reason for not stating them.

Coverage of all the published and unpublished information within the field it claims to cover is suggested as a standard to aim for. *Chemical Abstracts* offers the best example of thoroughness in this respect, but *Nuclear Science Abstracts* is very realistic in its coverage of both published and unpublished material. *Chemical Abstracts* issues every five years a list of journals and other serial material covered, with their locations in libraries. Photostat copies of articles may be obtained by the members of the American Chemical Society through a special service. *Nuclear Science Abstracts* calls attention to the system of depository libraries set up by the U.S. Atomic Energy Commission for its reports, and releases quantities of reports to the U.S. Office of Technical Services for general sale. Both journals indicate sources and prices for out of the way material, such as dissertations and reports. Staff-produced publications, such as the *Aeronautical Engineering Review* and the *Engineering Index,* are based upon libraries maintained for the use of members of the supporting society or the cooperating societies.

The *abstracts* should be analytical, describing theoretical or experimental parameters, results in physical, numerical, or chemical terms, methods and apparatus when new, references in the same abstract journal to closely related work, conclusions, and including an evaluation by the abstractor. The abstracts would combine, therefore, the best features of the abstracts in *Chemical Abstracts, Applied Mechanics Reviews,* the *Technical Abstract Bulletin (ASTIA),* and *Mathematical Reviews*. All abstracts would be signed. Frequently published lists of abstractors are desirable.

Subject indexes should be issued semi-annually *(Nuclear Science Abstracts)* and cumulated annually *(Applied Mechanics Reviews; Chemical Abstracts; Electrical Engineering Abstracts; Mathematical Reviews; Nuclear Science Abstracts; Physics Abstracts)*. Cumulated subject indexes for suitable periods are desirable (Decennial indexes of *Chemical Abstracts*).

Author indexes should be supplied with each issue *(Applied Mechanics Reviews, Chemical Abstracts, Electrical Engineering Abstracts, Mathematical Reviews, Nuclear Science Abstracts, Physics Abstracts)* and annually. It is desirable to cumulate author indexes also semi-annually *(Nuclear Science Abstracts)* and at intervals of several years (decennial indexes of *Chemical Abstracts*). It is especially desirable to include the titles of papers in each author index, as is done by *Chemical Abstracts, Electrical Engineering Abstracts, Mathematical Reviews,* and *Physics Abstracts*.

Corporate author or *report source indexes,* cumulated in the same manner as personal name author indexes, are desirable when required by the material covered, as is done by *Nuclear Science Abstracts* and as is announced by the *Technical Abstract Bulletin*.

Auxiliary indexes, additional to the usual author and subject indexes, are governed by the kinds of material covered by the abstract journal. For example, *Chemical Abstracts* added a formula index in 1920, and a numerical patent number index in 1935. *Nuclear Science Abstracts* includes numerical indexes of report numbers. U.S. *Government Research Reports* includes indexes of PB numbers, which are accession or report ordering numbers, and of Atomic Energy Commission report numbers. *The Technical Abstract Bulletin* includes indexes of AD accession or report ordering numbers.

The *recommended frequency* for an abstract journal is semi-monthly *(Chemical Abstracts, Nuclear Science Abstracts),* rather than monthly.

An *internal arrangement* of abstracts according to a detailed classification system is most helpful. *Physics Abstracts* and *Electrical Engineering Abstracts* follow this practice, using the Universal Decimal System. *Chemical Abstracts* has good internal arrangement of like materials.

Bibliographical form should include, for journals, the full title of the article, the author or authors, the title of the journal in a standard recognizable abbreviation, volume, number or date of issue, year, and inclusive pages *(Electrical Engineering Abstracts; Physics Abstracts; Nuclear Science Abstracts; Applied Mechanics Reviews; Engineering Index)*. *Chemical Abstracts* omits the month of issue. The *Aeronautical Engineering Review* has varied in both abbreviations and the inclusion of volume numbers and inclusive pages. At present, volume numbers are omitted. For reports, the bibliographical description should include the title, author or authors, issuing agency, series, report number, date of issue, and pages. Abbreviations of the names of agencies or series should be easily recognizable if abbreviated.

All *abstracts should be numbered,* for ease of indexing and finding. However, *Chemical Abstracts* has an excellent system of reference to specific fractions of each column for ready finding.

REFERENCES

1. BINNINGTON, JOHN P. The Integration of Report Literature and Journals. *American Documentation,* v. 3, August 1952, pp. 149–152.
2. CRANE, E.J., PATTERSON, AUSTIN M., and MARR, ELEANOR B. A Guide to the Literature of Chemistry. 2nd ed. New York, John Wiley & Sons, 1957. Chapter 10. Procedure in Literature Searches.
3. EVANS, ELMA. Reference Use of the Collection and Its Catalogs. *American Documentation,* v. 3, August 1952, pp. 146–149.
4. FRY, BERNARD M. Library Organization and Management of Technical Reports Literature. Washington, D.C., Catholic University of America Press, 1953. 140 pp.
5. GRAY, DWIGHT E. General Problems of Information Services. *Special Libraries,* v. 48, September 1957, pp. 313–320.
6. HERNER, SAUL. Information-Gathering Habits of Workers in Pure and Applied Science. *Industrial and Engineering Chemistry,* v. 46, January 1954, pp. 228–236.
7. LOWRY, W.KENNETH. Trends in United States Documentation Research. *Special Libraries,* v. 48, October 1948, pp. 364–366.
8. MILEK, JOHN T. Abstracting and Indexing Services in Electronics and Related Electrical Fields. *American Documentation,* v. 8, January 1957, pp. 5–21.
9. SMITH, MAURICE H. The James Forrestal Research Center Library. *The Princeton University Library Chronicle,* v. 16, No. 1, Autumn 1954, pp. 23–28.
10. TOULOUKIAN, Y.S. A Thermophysical Properties Research Center for the Collection, Correlation and Dissemination of Data and Information. *American Rocket Society. Preprint* No. 469–57, August 26–28, 1957. 27 pp.
11. VAN ANTWERPEN, F.J. Professional Societies and Professional Standards. *Special Libraries ,* v. 48, July–August, 1957, pp. 251–254.

PRINCIPAL ABSTRACT JOURNALS AND INDEXES DISCUSSED

Aeronautical Engineering Index. 1947 to date.

New York, Institute of the Aeronautical Sciences.

Aeronautical Engineering Review. 1942 to date.

New York, Institute of the Aeronautical Sciences.

Applied Mechanics Reviews. 1948 to date.

New York, American Society of Mechanical Engineers.

Chemical Abstracts. 1907 to date.

Columbus, Ohio, American Chemical Society.

Electrical Engineering Abstracts. (Section B of *Science Abstracts.*) 1903 to date.

London, Institution of Electrical Engineers.

Engineering Index. 1884 to date.

New York, Engineering Index, Inc.

Mathematical Reviews. 1939 to date.

Providence, R.I., American Mathematical Society.

Nuclear Science Abstracts. 1948 to date.

(U.S. Atomic Energy Commission.) Washington, D.C., U.S. Government Printing Office.

Physics Abstracts. (Section A of *Science Abstracts.*) 1903 to date.

London, Institution of Electrical Engineers.

Index of NACA Technical Publications. 1915–1949, 1949–June 1951, May 1951–June 1953, and annually.

Washington, D.C., U.S. National Advisory Committee for Aeronautics.

NACA Research Abstracts. June 15, 1951 to date.

Washington, D.C., U.S. National Advisory Committee for Aeronautics.

Technical Abstract Bulletin. September 1, 1957 to date.

Arlington, Va., U.S. Armed Services Technical Information Agency.

Technical Information Pilot. 1948–June 2, 1953.

Washington, D.C., U.S. Library of Congress, Technical Information Division, Navy Research Section. No more published.

Title Announcement Bulletin. March 13, 1953–July 1, 1957.

Dayton, Ohio, U.S. Armed Services Technical Information Agency. No more published.

U.S. Government Research Reports. 1946 to date.

Washington, D.C., U.S. Office of Technical Services (Department of Commerce).

APPENDIX. CASE HISTORIES

CHECKLIST OF CASES

1. Bibliography on the Origin of Turbulence and Related Subjects
2. Bibliography on Boundary Layer Control
3. Boundary Layer Control. Supplement
4. Bibliography on Helicopter Handling Qualities, Stability and Control, and Performance
5. Bibliography on VTOL and STOL Aircraft
6. Bibliography on the Atomic Rocket and Nuclear Propulsion
7. Liquid Ozone-Oxygen Mixtures as Rocket Oxidizers
8. Heat Transfer in Oscillating Flow
9. Mechanics and Kinetics of the Water-Gas Reaction
10. Wind Tunnel and Flight Testing of Rotors, Especially for Helicopter Stability
11. Taylor Instability
12. Helicopter Instrument Flying
13. Automatic Pilots for Helicopters
14. Influence on Lift of Vortex Shedding by Long Slender Bodies in Subsonic Flow
15. Drag of Submerged or Semi-Submerged Bodies in Water at Low Speeds

16. Flight Mechanics
17. Combustion of Fuel Sprays
18. Calculation of Induced Velocities (Downwash) of Supersonic Wings in Steady Flow
19. Calculation of Pressure Distribution on Sweptback Wings at Subsonic Speeds
20. Radial Inflow Turbines
21. Turboprops
22. Detonation in Gases and in Liquid and Solid Propellants
23. Poisoning of Combustion Catalysts
24. Automatic Temperature Controls for a Laboratory Furnace
25. Correlation Analyzers (Apparatus Only)
26. Power Spectrum Analysis
27. Reduction of Noise in the Reception of Pulse Signals by Auto-Correlation Techniques
28. Tables of Kelvin Functions for a Wide Range of Frequencies
29. Low-Frequency High-Powered Radio Transmitter
30. Mechanical Problems of Heavy Coil Windings
31. Properties of Kovar and Fernico
32. Colored Smoke Production
33. Ram-Jet Combustion and Turbojet Combustion
34. Thrust Reversal
35. Inertial Navigation
36. Optical Tooling
37. Dielectric Properties, Decomposition Products, and Toxicity of Sulfur Hexafluoride
38. Kinetics of Dissociation of CO_2
39. Air Pollution from Incomplete Combustion of Hydrocarbons
40. Shock Tube as a Chemical Research Tool
41. Sonic Flow in Small Orifices
42. Scintillation Counters—Design and Applications
43. Torquing of Stainless Steel Bolts
44. Welding of Stainless Steels
45. High-Temperature Wrap-Around Heat Insulating Material
46. Ferrites
47. Electroforming
48. Speech-Band Compression
49. Magnetic Clutches for Servomechanisms
50. Carrier Current Amplifiers

1. *Bibliography on the Origin of Turbulence and Related Subjects.* Literature Search No. 13, April 1957. 126 pp.

Source	Reports	Journals	Conferences	Total
Applied Mechanics Reviews	6	24	4	34
ASTIA Title Announcement Bulletin	17			17
ASTIA abstract cards	16			16
Physics Abstracts		13		13
NACA Research Abstracts	7			7
Chemical Abstracts		7		7
Brown University, Summaries on Compressible Flow	6			6
Engineering Index		2		2
Nuclear Science Abstracts	1			1
Card catalogs	2	2		4
Total	55	48	4	107
From the literature	35	51	6	92
Direct examination	22	20	43	85
Total	57	71	49	177
Unaccounted for	8	15	2	25
Books, dissertations, films	20			20
Total				329

This search was requested by a chemical engineering professor, and its undertaking was supported by two aeronautical engineering professors. Newly discovered information was sent to the requesting person as the search progressed, and there was considerable contact with the interested group during the course of the search, with beneficial exchanges.

2. *Bibliography on Boundary Layer Control.* Literature Search No. 6, January 14, 1955. 111 pp.

The basis for this bibliography was a collection of reports and reprints in the files of the requesting aeronautical engineering professor, and material in the Library. No existing bibliography covered the literature in the manner desired, especially for the period 1950–1954. Personnel of the research project in this field, as well as the directing professor, were in continual touch throughout this compilation, supplying references, advising on specific problems and encouraging its rapid completion. References were included from 12 bibliographies. The results of related searches on the electric analogy, the electrolytic tank, injection pumps and experimental techniques, especially flow visualization, were incorporated into the bibliography. Of 311 references included, 224 were reports, 72 were from journals, 9 were from congresses, and 2 each were books, preprints and dissertations.

Systematic searches were made of *Applied Mechanics Reviews,* 1950–Oct. 1954; *Engineering Index,* 1950–1953; *Physics Abstracts,* 1950–Oct. 1954; *Mathematical Reviews,* 1950–Oct. 1954; *Aeronautical Engineering Index,* 1950–1953; *Technical Information Pilot* (Library of Congress, Technical Information Division), 1950–June 2, 1953; *Title Announcement Bulletin* (Armed Services Technical Information Agency), March 13, 1953–Oct. 1954; and *Index of NACA Technical Publications,* 1915–May 1954. Records were not kept of the sources from which specific references were taken, but each of the sources searched produced something. *Applied Mechanics Reviews* was especially useful for foreign publications. The *Monthly List of Russian Accessions* (Library of Congress) was searched with negative results.

3. *Boundary Layer Control.* Supplement. On Cards, In Progress, January 1958.

Source	*Reports*	*Journals*	*Conferences*	*Total*
NACA Research Abstracts	135			135
Applied Mechanics Reviews	15	19		34
ASTIA Title Announcement Bulletin	12			12
Pacific Aeronautical Index '55		11		11
NACA Indexes	7			7
Physics Abstracts		3		3
Aeronautical Engineering Review		2		2
Chemical Abstracts		1		1
Battelle Technical Review		1		1
U.S. Government Research Reports	7			7
Total	176	37		213
NACA-ONR bibliography on cards	42			42
Direct examination	26	7	36	69
From the literature	11	7	1	19
ASTIA microcards	5			5
Card catalog	2			2
Total	86	14	37	137

4. *Bibliography on Helicopter Handling Qualities, Stability and Control, and Performance,* July 1953–February 15, 1956. Literature Search No. 10, February 15, 1956. 22 pp.

Types of Literature					
Reports	*Preprints*	*Journals*	*Congresses, etc.*	*Books*	*Total*
115	5	43	24	1	188

SOURCES USED

Title Announcement Bulletin (Armed Services Technical Information Agency), July 3, 1953–February 3, 1956.

NACA Research Abstracts (National Advisory Committee for Aeronautics), July 1953–January 1956.

Index of NACA Technical Publications, June 1953–May 1955.

U.S. Government Research Reports, July 1953–January 1956.

Applied Mechanics Reviews, January 1954–January 1956.

Engineering Index, 1953 and 1954.

Aeronautical Engineering Review, January 1954–January 1956.

United Aircraft Corporation. Library. Index to Current Technical Publications, January 3, 1955–February 13, 1956.

Princeton University. Forrestal Research Center. Library. Acquisitions, January 5, 1955– February 5, 1956.

Also the Library card catalog; abstract cards from the Armed Services Technical Information Agency, and microcards from the same source.

This search was requested by an aeronautical engineering professor. The original aim was to present all the new information available on helicopter handling qualities, taking up where existing bibliographies left off. This was promised for February 15, 1956, just three weeks from the date of the request, and was delivered on that day.

Discussion during the search with other research group personnel working in the field of helicopter stability and control led to a broadening of the scope of the search, as the title indicates. The time limitation precluded more than the bibliographical description of each item. No abstracts could be included.

Time spent on the search was not measured. However, one of the three weeks was taken up with typing and printing, leaving two weeks in which to compile the search, or approximately 80 working hours. It is doubtful if as much as 40 hours of compilation time were required.

5. *Bibliography on VTOL and STOL Aircraft.* Literature Search No. 12, January 1957. 56 pp.

Source	*Reports*	*Journals*	*Conferences*	*Total*
ASTIA Title Announcement Bulletin	29			29
NACA Research Abstracts	18			18
Aeronautical Engineering Review	1	13		14
Pacific Aeronautical Index '55		13		13
Applied Mechanics Reviews	5	6		11
Forrestal Research Center Acquisitions List		4		4
United Aircraft Corp. Index	2	4		6
Battelle Technical Review		1		1
Engineering Index		1		1
Total	55	42		97
Direct examination	23	13	15	51
From the literature	30	2	—	32
Total	53	15	15	83

This search was requested by an aeronautical engineering professor on behalf of a research group. Trade journal descriptive material on specific aircraft was included. References unaccounted for in the table were largely identified through ASTIA abstract cards, ASTIA microcards, and library card catalogs.

6. *Bibliography on the Atomic Rocket and Nuclear Propulsion.* Literature Search No. 11, February 10, 1957. 33 pp.

Source	*Reports*	*Journals*	*Conferences*	*Total*
Nuclear Science Abstracts	14	6		20
Aeronautical Engineering Index		17		17
Engineering Index		5		5
United Aircraft Corp. Index		5		5
NACA Indexes	3			3
ASTIA Title Announcement Bulletin	3			3
Chemical Abstracts		2		2
Aeronautical Engineering Review	1			1
NACA Research Abstracts	1			1
Forrestal Research Center, Acquisitions List		1		1
Total	22	36		58
Direct examination	2	18	9	29
From the literature	1	10	2	13
Total	3	28	11	42

Two aeronautical engineering professors and the chief engineer of the jet propulsion research group requested this search. While original research papers were emphasized, the bibliography includes some trade journal material and some reports on equipment.

7. *Liquid Ozone-Oxygen Mixtures as Rocket Oxidizers.* On cards. In progress, January 1958.

Source	*Books*	*Journals*	*Reports*	*Congresses*	*Patents*	*Dissertations*	*Total*
Chemical Abstracts		34		1	6	2	43
Physics Abstracts		2					2
Total		36		1	6	2	45
From the literature	1	5	1			1	8
Direct examination	2	4	5	2	1		14
ASTIA visit			2				2
Meeting program				11			11
Forrestal Research Center Library Acquisitions		2					2
Title Announcement Bulletin (ASTIA)	1						1
Total	4	11	8	13	1	1	38

This will be a revised and expanded edition of Literature Search No. 8, May 20, 1955. It was requested originally by an aeronautical engineering professor, and the current work is done under the guidance of the chief engineer of jet propulsion projects.

8. *Heat Transfer in Oscillating Flow.* On cards. In progress, January 1958.

By arrangement with the project leader, an engineer of the project is examining and evaluating all material for this search. A previous literature search on this subject was issued in 1955 (Literature Search No. 7).

The new material available since that time, and some older material not in the earlier search, includes 19 items of direct interest. Of these items, one each was located through the *Engineering Index, Nuclear Science Abstracts, Physics Abstracts,* and *Dissertation Abstracts.* The remaining 15 items were derived from direct examination or from authors' citations.

Of 32 other evaluated titles to be included, *Chemical Abstracts* produced 1; *Engineering Index,* 4; *Nuclear Science Abstracts,* 3; *Applied Mechanics Reviews,* 2; *Physics Abstracts,* 5; *Dissertation Abstracts,* 2. The remaining 15 were located by direct examination or through authors' citations.

There remain about 150 journal papers and reports to be evaluated.

9. *Mechanics and Kinetics of the Water-Gas Reaction,* 1903–March 10, 1953. Literature Search No. 5, March 18, 1953. 12 pp.

This search was requested by a research chemist, in the words of the title. Some German work new to him was found through *Chemical Abstracts,* but the search did not progress further, because of pressures elsewhere.

A succeeding chemist brought references familiar and useful to him, rephrased the problem to include "the carbon monoxide steam reaction," and communicated frequently throughout until the search was finished.

Two doctoral dissertations and 25 journal references were discovered through *Chemical Abstracts;* and 34 journal references and 10 books were added from references cited in the literature, by direct examination or directly from the requesting chemist. Abstracts were provided for about half the items, partly in consultation with the chemist, partly on the judgment of the searcher.

10. *Wind Tunnel and Flight Testing of Rotors, Especially for Helicopter Stability.* Literature Search No. 3, October 27, 1952. 17 pp.

The period since 1945 was emphasized. The request was made by an engineer with a helicopter project. The sources searched systematically were the *Engineering Index, Aeronautical Engineering Index, Aeronautical Engineering Review, Applied Mechanics Reviews, NACA Research Abstracts,* and *Technical Information Pilot* (Library of Congress, Technical Information Division). The bibliography included 98 reports, 41 journal papers, and 16 papers in proceedings of congresses. No records were kept of the sources from which specific items were derived.

11. *Taylor Instability.* December 1957.

The point of entry into the literature was the card catalog under "Taylor Instability," locating Atomic Energy Commission reports of 1952 and later. From references in these, the original Taylor paper of 1950 was located. Subsequently, references were located, as follows:

Source	*Reports*	*Journals*	*Total*
Originally located in card catalog	9		9
Nuclear Science Abstracts, 1950 to date	4		4
Physics Abstracts, 1950 to date		2	2
Direct examination of current journal by the inquirer		1	1
From the literature	2	10	12
Total	15	13	28

One book (Lamb's *Hydrodynamics*) was also referenced in the literature. *Nuclear Science Abstracts* picked up certain AEC reports of 1952 and 1953 in 1956. The inquirer was a professor of aeronautical engineering. Estimated searching time, 3 hours. The abstract journals were searched last in this case.

12. *Helicopter Instrument Flying.* November 1957.

Source	*Reports*	*Journals*	*Conference proceedings*	*Total*
Engineering Index, 1950 to date	3	1		4
Aeronautical Engineering Index, 1950–1955		3		3
Aeronautical Engineering Review, 1956–Oct. 1957		3		3
Direct examination		2		2
From the literature		1		1
Card catalog			2	2
From previous brief search	3		3	6
Total	6	10	5	21

Estimated time, 2 hours. Effectiveness: no new information expected or found. Work discontinued at suggestion of inquirers. Requested by a helicopter test pilot and an electronics engineer. Work carried out simultaneously with search on Automatic Pilots for Helicopters.

13. *Automatic Pilots for Helicopters.* November 1957.

Source	*Reports*	*Journals*	*Conference proceedings*	*Total*
Engineering Index, 1950 to date		4		4
Air University Periodical Index		4		4
Aeronautical Engineering Index		3		3
Title Announcement Bulletin	1			1
Previous brief search	2	1	1	4
Direct examination		3		3
From card catalog			2	2
Total	3	15	3	21

Estimated time, 2 hours. Effectiveness: no new information expected or found. Work discontinued at suggestion of inquirers. Requested by a helicopter test pilot and an electronics engineer. Work carried out simultaneously with search on Helicopter Instrument Flying.

14. *Influence on Lift of Vortex Shedding by Long Slender Bodies in Subsonic Flow.* September 1957.

A paper in the *Journal of the Aeronautical Sciences* for June 1957, containing 18 references, and an unclassified paper in a classified conference volume of 1956, containing 25 references, were located by direct examination. These were considered to have essentially the bibliography. However, *NACA Technical Note* 1662, in the hands of the requesting aeronautical engineering professor, turned up two additional journal references. Estimated time, about 3 hours. Work discontinued at the suggestion of the requestor. Of approximately 10 references pertaining to the subsonic regime located by the means described, about 5 were used intensively. Only NACA indexes and the card catalog were scanned.

15. *Drag of Submerged or Semi-Submerged Bodies in Water at Low Speeds.* April 1956.

The point of entry into the literature was through direct examination of publications of the David W.Taylor Model Basin and of articles in the *Aeronautical Engineering Review.* These turned up 3 reports and one article, which produced much of the bibliography. James L.G.FitzPatrick's *Bibliography on Flapping Wing Flight* (1950) produced 13 more references. Subsequent searches of the *Zoological Record* and *Biological Abstracts* for recent years turned up only known references. The requesting aeronautical engineering professor was provided with reprints of articles by another professor, and work was discontinued at the suggestion of the requestor. Estimated time, about 6 hours.

16. *Flight Mechanics.* October 6, 1957. 18 pp.

Two aeronautical engineering professors requested a bibliography on flight mechanics, to be representative rather than exhaustive, and to cover such subjects as

Orbits and Trajectories of Rockets and Missiles and Transfer between Orbits.

To accomplish this in a relatively short time, three journals were examined directly: *Astronautica Acta, Jet Propulsion,* and the *Journal of the British Astronautical Society*. In addition to these the available proceedings of International Astronautical Congresses were examined. The bibliography was built up principally from these sources, together with authors' references in them. Use was made also of about the last two years of the *Aeronautical Engineering Review,* the Abstract section of the *Journal of the British Interplanetary Society,* and the latest *Index of Publications* of the Rand Corporation. About 100 references were included, without abstracts.

At the request of the senior professor, a supplement was prepared, consisting largely of books on advanced dynamics and some astronautical titles.

17. *Combustion of Fuel Sprays.* Typewritten. July 20, 1954, 11 pp. July 22, 1954. 5 pp.

The requesting aeronautical engineering professor was interested in recent work, especially recently declassified material. The lists as submitted covered the period 1950 to date. Principal sources were *NACA Research Abstracts, Chemical Abstracts,* the Penn State *Bibliography on Sprays,* 2nd ed., January 1953, and the *Aeronautical Engineering Review,* together with various authors' citations. About 115 references were included.

18. *Calculation of Induced Velocities (Downwash) of Supersonic Wings in Steady Flow.* Typewritten. November 16, 1951. 2 pp. Supplement, November 19, 1951. 1 p.

An aeronautical engineering professor made this request. The two lists were compiled by direct examination of recent files of NACA, British and Douglas Aircraft Company reports and the *Journal of the Aeronautical Sciences*. References cited by various authors were added. The two lists included respectively 21 and 10 references.

19. *Calculation of Pressure Distribution on Sweptback Wings at Subsonic Speeds.* Typewritten. March 5, 1952. 2 pp.

An aeronautical engineering professor made this request. Several reports, both American (NACA) and British were assembled by going through the reports for the last two years, then checking through the authors' bibliographies. The complete typed list included 19 references.

20. *Radial Inflow Turbines.* Literature Search No. 1, September 17, 1952. 2 pp.

An aeronautical engineering professor requested this search, which had to be quick. About 20 references were included, largely from two papers by W.T.von der Nuell and O.E.Balje, both published in 1952. The *Engineering Index, Applied Mechanics Reviews, Aeronautical Engineering Index, Aeronautical Engineering Review,* and *NACA Research Abstracts* were used also.

21. *Turboprops.* Literature Search No. 2, September 18, 1952. 6 pp.

The requesting aeronautical engineering professor was interested in the basic problems to be solved. Descriptive data were not of interest. About 55 readily available reports and journal articles were located through the *Aeronautical Engineering*

Index, Aeronautical Engineering Review, Engineering Index, Applied Mechanics Reviews, and the National Bureau of Standards Circular 509 (1951), which is a bibliography by Fiock and Halpern.

22. *Detonation in Gases and in Liquid and Solid Propellants,* 1948 to 1953. (Completed February 13, 1953).

The bibliography was compiled at the request of an aeronautical engineering professor for a quick survey of the literature for the last five years. About 90 references were obtained through *Chemical Abstracts, Physics Abstracts, Applied Mechanics Reviews,* and the *Engineering Index.* No record of the sources which produced specific references was kept.

23. *Poisoning of Combustion Catalysts.* January 1957.

This inquiry came from an electronics engineer interested in the mechanism of the poisoning reaction. The literature was entered through some reports of the Hydrogen Peroxide Laboratory of the Massachusetts Institute of Technology, previously familiar, and by going through the annual review volumes, *Advances in Catalysis,* where the review paper by E.B.Maxted was found in the 1951 volume. From the bibliographies in these sources, the original work of Coward and Guest *(J. Amer. Chem. Soc.,* v. 49, Oct. 1927, pp. 2479–2486) and by Guest *(U.S. Bureau of Mines Technical Paper* No. 475, 1930), were found. AGARDograph No. 7, "Introduction to the study of chemical reactions in flow systems," by S.S.Penner (1955) was added. The inquirer selected the papers by Coward and Guest, Maxted, and one MIT report and later requested the 1930 Bureau of Mines paper by Guest. The searching time was slightly under one hour.

24. *Automatic Temperature Controls for a Laboratory Furnace.* 1952.

A list of 17 references was compiled from *Chemical Abstracts, Engineering Index, Physics Abstracts,* and *Nuclear Science Abstracts.* The investigator, a metallurgist, selected 8 of these and studied them. He was interested in designing a low-cost control. The professor in charge of the laboratory returned from out of town and decided they would buy a suitable control, as there was not time to design and build one. *Chemical Abstracts* and the *Engineering Index* were the most useful sources, each accounting for half of the references used.

25. *Correlation Analyzers* (Apparatus Only). June 1957.

Source	*Reports*	*Journals*	*Congresses, etc.*	*Theses*	*Total*
Physics Abstracts		5			5
Engineering Index	2	2			
Electrical Engineering Abstracts	3	3			
Total	10	10			
From the literature	3	6	2	4	15

The requesting person was an electronics engineer with a nuclear physics project. The object was to get a quick review of the known types of equipment. A typed list was submitted the following day, and the papers immediately available were assembled. Approximately three hours of compiling time were spent.

26. *Power Spectrum Analysis.* June 1957.

A list of 18 references was compiled and the material assembled for inspection by the inquiring electronics engineer. The literature was entered through direct examination of *NACA Technical Notes* by J.M.Eggleston, and the *AGARD Flight Test Manual;* and the bibliography was compiled from references cited by various authors. Three titles were of interest for immediate use: Wiener (1949), Tukey (1949), and Lee (1950).

27. *Reduction of Noise in the Reception of Pulse Signals by Auto-Correlation Techniques.* October 1957.

The literature was entered through the *Engineering Index* in the presence of the requesting electronics engineer, under such subjects as Radio circuits—noise; Radio equipment—noise; Radio detectors; and Radio communication. Subsequently 11 references to journals were located for the years 1955 and 1956, and about half of these were selected for use. Work was discontinued at the suggestion of the requestor. Estimated searching time, 1 1/2 hours.

28. *Tables of Kelvin Functions for a Wide Range of Frequencies.* June 1957.

This request, from an engineer with a nuclear physics project, was investigated through the Fletcher, Miller, and Rosenhead *Index of Mathematical Tables; Mathematical Tables and Other Aids to Computation,* v.1, 1943 to v.11, January 1957; *Mathematical Reviews,* 1954-March 1957; and Davis, *Bibliography and Index of Mathematical Tables,* 1949. Results beyond those recorded in Bateman and Archibald's “Guide to Tables of Bessel Functions,” *MTAC,* v.1, July 1944, pp. 205–308, were few. Pressure of other work since the matter was investigated has prevented further search. However, in January 1958 one of the tables found, published in 1954 *(Bull. Soc. Roy. Sci. Liège,* v.23, pp. 52–59), was requested.

29. *Low-Frequency High-Powered Radio Transmitter.* 1955.

Of 7 references located through the *Engineering Index* and 4 through the *Electrical Engineering Abstracts* (1947 to date), the requesting electrical engineer selected 3 for study. The estimated searching time was 2 hours.

30. *Mechanical Problems of Heavy Coil Windings.* September 1957.

This was requested by an electrical engineer. While a bibliography of 18 references was obtained through the *Engineering Index,* 1947 through 1956, only one item was actually obtained and submitted: W.Querfurth, “Coil Winding,” Chicago, G. Stevens Manufacturing Co., 1954. It was considered adequate, together with information obtained on trips by the inquirer. Estimated time, 2 hours. Work discontinued with approval of the inquirer.

31. *Properties of Kovar and Fernico.* January 1956.

Some 14 items were collected for the use of the requesting engineer with a nuclear physics project, through *Physics Abstracts* and *Chemical Abstracts,* published between 1935 and December 1955. Data on the desired electrical and magnetic properties were not located. A manufacturer's bulletin obtained by the requestor was the best single reference.

32. *Colored Smoke Production.* September 18, 1953.

This search was conducted in *Chemical Abstracts,* 1935 to date. Abstracts were provided of the three journal articles and seven patents turned up. The inquiring aeronautical engineering professor was not interested in more than these.

33. *Ram-Jet Combustion and Turbojet Combustion.* February 6, 1953. 3 pp.

An aeronautical engineering professor requested a quick three-year survey of the principal contributions in these two fields. The sources used were the *Aeronautical Engineering Review, Aeronautical Engineering Index,* and the *Engineering Index.* Recent NACA and British reports were examined directly. Authors' citations were used in the usual manner. 12 references on ramjet combustion and 17 on turbojet combustion were submitted.

34. *Thrust Reversal.* April 1954.

Material for this inquiry by an aeronautical engineering professor was located in recent issues of aeronautical news magazines and by use of the *Aeronautical Engineering Review,* the United Aircraft Corporation weekly index, and the Air University *Periodical Index.* Eight articles were assembled.

35. *Inertial Navigation.* October 1956.

The inquiring person was a civil engineering professor. The principal source was the *Aeronautical Engineering Review*. The inquirer wished to see all of the 12 articles listed.

36. *Optical Tooling.* November 1956.

The inquiring person was a civil engineering professor. The principal sources were the *Aeronautical Engineering Review* and the *Engineering Index* and a visiting lecturer in aeronautical engineering. Of 10 articles located, the inquirer wished to see all.

37. *Dielectric Properties, Decomposition Products, and Toxicity of Sulfur Hexafluoride.* April 15, 1955.

A list of 24 references was submitted to the inquiring engineer with a nuclear physics project. 6 of the immediately available papers were selected by him for study. *Chemical Abstracts* was the only source used, and only to the end of 1953. The search was completed by the inquiring engineer for more recent publications.

38. *Kinetics of Dissociation of* CO_2. August 1957.

Seven published papers were examined. One of these *(Journal of Chemical Physics,* v.23, May 1955, pp. 902–908), based upon a doctoral dissertation, contained the mechanism of the dissociation, and a continuation (same journal, v.26, June 1957, pp. 1727–1733) contained a confirmation of the mechanism. The inquiring chemist accepted these and desired no other literature. *Chemical Abstracts* was used.

39. *Air Pollution from Incomplete Combustion of Hydrocarbons.* October 1957.

The inquiring chemist (Ph.D.) was “familiar with chemical sources, not with others.” While he selected five references to 3 journals and 2 international conferences from the *Engineering Index,* several articles were located by direct examination of the *SAE Transactions* for 1955, and the article on "Smokes and Fumes" in the *Encyclopedia of Chemical Technology*. Work was discontinued after subsequent location of information in current issues of the *SAE Journal*.

40. *Shock Tube as a Chemical Research Tool.* January 1953.

This request came from an aeronautical engineering professor. Inquiry was made of a physical chemist, who confirmed that little was available on the subject except a recent letter to the editor in *the Journal of Chemical Physics* (v.19, Oct. 1951, p. 1313), and possibly a recent paper in the *Journal of Applied Physics* (v.23, Dec. 1952, pp. 1390–1399).

41. *Sonic Flow in Small Orifices.* 1954.

Actually, the two papers which satisfied this inquiry by a chemical engineer were sufficiently identified by a chemist who had mentioned them. Both were published in the *Review of Scientific Instruments* (v.20, Jan. 1949, pp. 61–66, and v.21, Jan. 1950, pp. 25–30). Other references supplied were not of interest, as the orifices were too large.

42. *Scintillation Counters—Design and Applications.* October 29, 1957.

For this inquiry from an engineer with a nuclear physics project, two recent books were supplied, and pages were copied by Verifax from the *Engineering Index,* 1953–1956. These were sufficient for the time being.

43. *Torquing of Stainless Steel Bolts.* 1956.

An electrical engineer with a nuclear energy project requested this information. A few aeronautical sources supplied some immediate information (*Handbook of Instructions for Aircraft Designers; SAE Special Publication* No. SP-23), and three articles were located through the *Engineering Index.* The most satisfactory information was obtained from the Industrial Fasteners Institute: "Torquing of Non-ferrous and Stainless Steel Bolts," *Fasteners,* v.9, No. 5, 1954, pp. 3–7, also located through the *Engineering Index.* About one hour was needed for this search.

44. *Welding of Stainless Steels.* 1956.

The request came from an engineer with a nuclear physics project. The *Engineering Index* produced 6 likely references in a quick search over the years 1952–1955. Two of these, a handbook on welding and a book on stainless steels were used.

45. *High-Temperature Wrap-Around Heat Insulating Material.* 1955.

One reference was obtained by the requesting electrical engineer through the *Engineering Index* while the searcher located two Atomic Energy Commission Reports through *Nuclear Science Abstracts*. The engineer said he found information on whom to get in touch with through these three references, and work was discontinued.

46. *Ferrites.* (General Information) 1955.

This inquiry came from an electrical engineer with a nuclear physics project. Pages of the *Engineering Index* listing the literature for three years were copied by

Verifax and sent to him, and he selected two or three articles which supplied the information. A metallurgist was supplied with a bibliography in the same way at his request.

47. *Electroforming.* August 1957.

The requesting engineer is with a nuclear physics project. Five articles were assembled, and two books were obtained. The principal sources were the *Engineering Index* and library card catalogs.

48. *Speech-Band Compression.* January 1958.

Requested by an electronics engineer, to cover only literature published since January 1957, on electronic apparatus. Principal source, suggested by the requestor, was the *Journal of the Acoustical Society of America.* Seven titles of abstracts and one article were recorded. Five of the abstracts of papers presented and one patent were of interest. These revealed work by two industrial laboratories, two universities, and one Air Force laboratory, under one Army and two Air Force contracts. A subsequent search of *Physics Abstracts* for the year 1957 turned up no additional material. This search was conducted within 15 minutes, as *Physics Abstracts* arranges material under Vibration and Acoustics according to the universal decimal system, and all pertinent abstracts to the question would bear the numbers 534.78 or 534.781. The search *of the Journal of the Acoustical Society of America* required 30 minutes, with the index in the December 1957 issue. Work was discontinued at the suggestion of the inquirer.

49. *Magnetic Clutches for Servomechanisms.* January 1958.

The requesting electronics engineer defined the field exactly, having first asked for a few articles on magnetic clutch applications. He had checked in some books without results. Sufficient information was found in two articles *(Electronics,* v.22, Nov. 1949, pp. 100–103; and *Tele-Tech,* v.11, Sept. 1952, pp. 90 plus) through the *Engineering Index* under Clutches, magnetic. It will be noted that better results for this particular purpose were obtained at dates near the date of the first appearance of the magnetic clutch (1948).

50. *Carrier Current Amplifiers.* September 8, 1953.

Five references were submitted to the inquiring electronics engineer, all from *Electrical Engineering Abstracts,* 1946 to date. The best information was found later, in British R & M No. 2627, September 1947, “Electronics Applied to the Measurement of Physical Quantities.”

Analytical Study of a Method for Literature Search in Abstracting Journals

P.S.LYKOUDIS, P.E.LILEY, and Y.S.TOULOUKIAN

ABSTRACT. The mathematical analysis of a method for literature search, using abstracting journals, is presented based on the following model. One starts searching the abstracting journal beginning with the most recent issue and going back through a number of years, *a*. Next, the bibliographies of the papers located in these *a* years are searched for new references. The references found in this second step of the search will, in general, cover a period of years, *(b–a)*. Then one reverts back to searching through the abstracting journal again for another interval of *a* years starting with the year *b*. This cyclic procedure of alternate searches through the abstracting journal, followed by searching the bibliographies of uncovered papers is repeated until the total number of desired years of search is covered.

The cost of search by the above method is compared with the cost of the conventional method of search by simply going through all volumes of the abstracting journal. The ratio of these two costs is investigated. Equations are given for the evaluation of the search parameters *a* and *b* based on a study of the completeness of coverage of the period *(b–a)* years through the search of bibliographies. This part of the analysis is informative in itself since it determines the average number of years covered by the references contained in the bibliographies of papers located in a period of *a* years of searching through the abstracting journal.

An illustrative example is presented to demonstrate the use of the results obtained from the analysis.

1. INTRODUCTION

The world's technical literature constitutes a wealth of information representing the record of experiences and technical achievements of thousands of scientists and engineers. During the last decade, the volume of this literature has been increasing at such a rapid rate that it has become nearly impractical to bring it under absolute control. On the other hand, in this period of highly

P.S.LYKOUDIS, P.E.LILEY, and Y.S.TOULOUKIAN Thermophysical Properties Research Center, School of Mechanical Engineering, Purdue University, Lafayette, Indiana.

accelerated scientific and technological developments, the need for keeping abreast of the world's technical literature in a given field has become an absolute necessity in order to conserve such valuable resources as time, technical manpower, and money.

Recognizing the seriousness of this situation in the field of thermal properties of materials, the Thermophysical Properties Research Center (TPRC) was established in January 1957. The aims, functions, and operations of this Center are discussed in some detail in the Center's prospectus, TPRC Pub. A-2. As part of its over-all research activities the Thermophysical Properties Research Center searches the world literature for the thermophysical properties of gases, liquids, and solids, and compiles this information for use in its own research work as well as for dissemination at large.

Information pertinent to the interests of TPRC comes to the attention of the Center through many channels. These channels embrace the abstracting journals in pure and applied sciences, listings of special governmental and industrial technical reports, academic work reported in dissertations and theses, major compendia, and other more fugitive sources. Among the various sources of information enumerated above, abstracting journals constitute the most extensive and readily accessible source of literature references. Since many of the major abstracting journals have been issued over a period of several years; they provide a rather unique source for a systematic search of the literature.

A brief familiarization with any one of the major abstracting journals will enable one to realize that a brute force, page by page, scanning of the issues of abstracting journals is prohibitive in time and effort. Furthermore, since the search of such journals must often be performed by rather highly trained scientific and technical personnel, the cost of such a search effort is very great. In view of these factors, and considering the broad scope of operations undertaken by TPRC, an effort was made to develop an efficient procedure for searching the technical literature in abstracting journals. Certain phases of the study reported herein have been in use at TPRC for nearly a year with extremely rewarding results. It is to be realized that the total aspect of the study presented herein embodies certain simplifying assumptions and that certain of the considerations may need modifications as a result of further experience. It should be emphasized that experimental data are needed to substantiate the details of the proposed method fully.

The authors wish to express their sincere appreciation to Messrs. T.Wing and R.Rodine, of the TPRC staff, for their careful study of the manuscript and their many valuable suggestions. Mr. Wing also kindly prepared all the figures in this paper.

2. GENERAL CONSIDERATIONS

Consider the method depicted by Fig. 1 for searching a set of volumes of an abstracting journal. This is a schematic diagram of the percentage yield, p, plotted as a function of age, x, of issues surveyed, the search starting at the most recent year available, x=0. Obviously, such a representation can be obtained only for subjects which have received review in the abstracting journal for a period of years. Consequently, the search method described below will not be applicable to the literature of recent developments such as magnetohydrodynamics. A detailed analysis of the cycle in Fig. 1 is given in Sec 5.

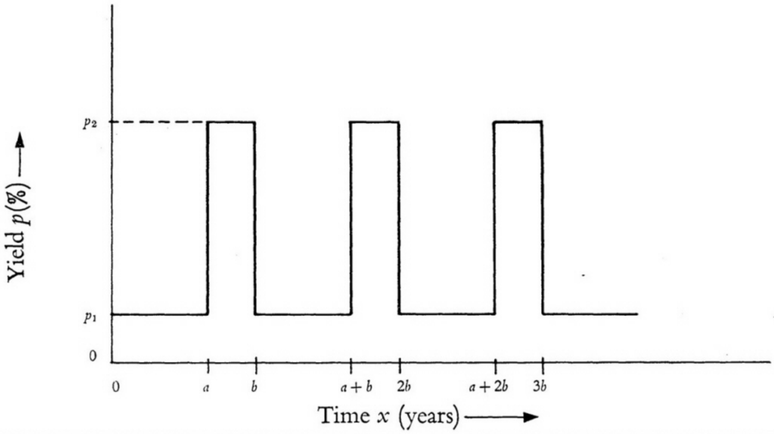


FIGURE 1. Search cycle treated in analysis.

During the first a years (starting with the most recent year) the abstracts are searched giving a yield, say p_1. This number is defined as the per cent ratio of useful abstracts found (those pertinent to the subject matter under investigation) to the total number of abstracts searched. The following $(b-a)$ years will be covered by searching the references listed in the papers located during the abstract journal search of the first a years. Let the yield for this period be p_2. The search will be completed by repeating this cycle an appropriate number of times as indicated in Fig. 1.

Consider an abstracting journal which goes back a number of years, Y. Should the search be completed in an integral number of cycles, then

$$Y = n_1 a + n_1 (b - a) = n_1 b \tag{1}$$

On the other hand, should the search be terminated with the abstracting journal, thus resulting in a non-integral number of cycles, one obtains

$$Y=(n_2+1)a+n_2(b-a)=a+n_2b \tag{2}$$

where n_1 and n_2 are appropriate integers.

The cost of search, C_1, by the method leading to Equation (1) is given by

$$C_1=K\,[n_1a+n_1\,(b-a)\,(p_1/p_2)] \tag{3}$$

while the cost, C_2, by the method leading to Equation (2) is similarly given by

$$C_2=K\,[(n_2+1)a+n_2(b-a)\,(p_1/p_2)] \tag{4}$$

where K in both equations is a proportionality factor.

Equations (3) and (4) are based on the assumption that the number of references located through the search of the papers collected during the a years is equal to the number of references that would have been located if one had chosen to cover the period $(b-a)$ years by searching through the abstracting journal.

The conventional procedure of searching in an abstracting journal is by going through the abstracts for all the years, Y. Let the cost of such a search procedure, which will be proportional to Y, be designated as C_a. With these definitions of costs one can form the following ratios:

$$R_1=\frac{C_1}{C_a}=\frac{n_1a+n_1\,(b-a)\,(p_1/p_2)}{n_1a+n_1\,(b-a)}=\frac{a}{b}+\left(1-\frac{a}{b}\right)\frac{p_1}{p_2} \tag{5}$$

$$R_2=\frac{C_2}{C_a}=\frac{(n_2+1)\,a+n_2\,(b-a)\,(p_1/p_2)}{(n_2+1)\,a+n_2\,(b-a)}=\frac{(1+1/n_2)\cdot[a/(b-a)]+(p_1/p_2)}{(1+1/n_2)\cdot[a/(b-a)]+1} \tag{6}$$ [1]

In the following sections the determination of the parameters, a, b, p_1 and p_2 is discussed.

3 EVALUATION OF THE YIELD P1

To establish the value of p_1 it is assumed that the information contained in all the issues of one complete year of the abstracting journal is adequate.[2] Let the total number of abstracts contained in a given volume of an abstracting journal be equal or proportional to E_t. Let the total number of useful abstracts located

[1] Should n_2 be a large number compared to unity, then the ratio R_2 approaches R_1.

[2] This is true for most abstracting journals for periods of about five years. In certain cases it is advisable to make a complete volume search, say every five years, to check on possible variations

by going through all the entries of this volume be E_u. The individual issues of an abstracting journal are divided into a number of sections. Let the total number of these sections be denoted by *s* and the total number of abstracts contained in the *i*th section and the corresponding number of useful abstracts be $(E_t)_i$ and $(E_u)_i$, respectively. With these notations, the following ratios are defined:

$$(Q_t)_i = \frac{(E_t)_i}{E_t} \tag{7}$$

$$(Q_u)_i = \frac{(Q_u)_i}{Q_u} \tag{8}$$

The quantities $(Q_t)_i$ and $(Q_u)_i$ are calculated and then ordered in such a manner that

$$(Q_u)_{i+1} \geqq (Q_u)_i \tag{9}$$

Now it is possible to form the following s partial sums:

$$(T_u)_k = \sum_{i=1}^{k} (Q_u)_i \qquad (k = 1, 2, \cdots, s) \tag{10}$$

and

$$(T_t)_k = \sum_{i=1}^{k} (Q_t)_i \qquad (k = 1, 2, \cdots, s) \tag{11}$$

It is obvious that, owing to the inequality (9), Equation (10) yields the relations:

$$(T_u)_{k+1} \geqq (T_u)_k$$

and

$$(T_u)_{k+1} - (T_u)_k \geqq (T_u)_k - (T_u)_{k-1} \tag{12}$$

Equations similar to Equation (12) are not possible for the sum given by Equation (11).

Now the pairs of corresponding points $(T_u)_k$ and $(T_t)_k$, which are *s* in number, can be represented in graphical form as shown by curve *A* of Fig. 2[3]. From this figure it is seen that a number of *s* distinct points are obtained. For cases where a large number of sections are surveyed these points will lie on a continuous curve in the interval from k=1 to k=s. Curve *A* gives useful information on the distribution of abstracts of interest contained in the various sections of the abstracting journal.

In Fig. 2 three distinct types of distribution curves are shown. All three curves (*A, B,* and *C*) refer to the same subject matter obtained from three different abstracting journals.

[3] This curve is representative of the type of subject investigated in Sec. 7.

The distribution curve, *A,* is very steep at the origin showing that a high useful yield is obtained by searching only a small fraction of the complete issue of the abstracting journal. Also, for high values of $(T_t)_k$ the gain in the yield $(T_u)_k$ is very small. This trend indicates that most of the useful information is found concentrated in a few sections containing altogether a rather small proportion of the total number of abstracts in the volume.

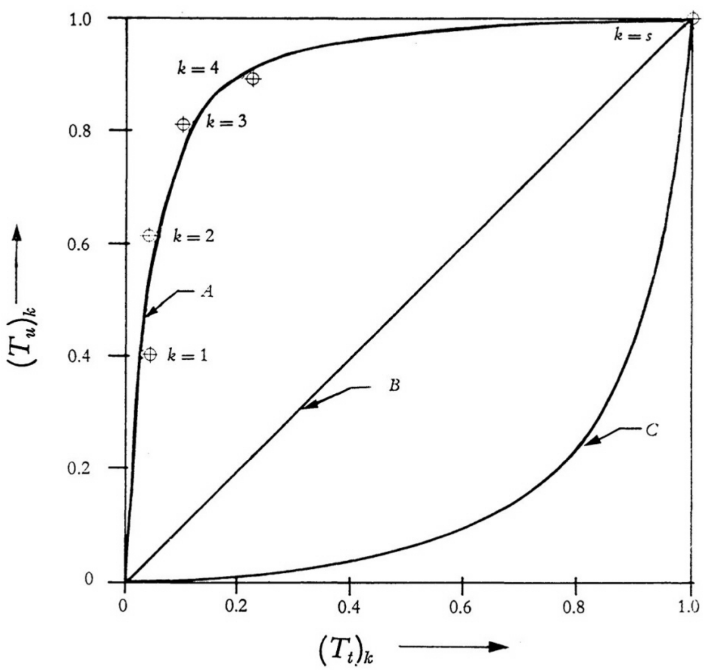


FIGURE 2. Typical yield distributions for different types of abstracting journals.

Curve *B* is a straight line showing that the useful information is evenly distributed among all sections. In this instance, the yield is directly proportional to the search effort.

Curve *C* shows the case of an abstracting journal where much search is needed for a small yield. In this type of abstracting journal one has to search most of the volume to redeem a satisfactory percentage of the total useful information.

It is obvious from the above analysis that an abstracting journal of type *A* is to be preferred to those of types *B* or *C*. Calculation of the slopes near the origin of the curves in Fig. 2 is useful since they are a measure of the cost for searching a particular abstracting journal. However, deviations from this criterion may be necessitated by the nature of the subject matter under study. For example, should a subject of interest be covered mainly in a section with an otherwise extremely low yield, one has to include this section in the search although other sections with higher yields may have been excluded. Conversely, one may exclude a section with a relatively high number of useful abstracts but with the ratio $(Q_u)_i/(Q_t)_i \ll 1$.

From the above remarks one is led to conclude that an analysis of the ratio $(Q_u)_i/(Q_t)_i$ gives useful guidance in the final selection of the specific sections to be searched.

Having decided on the optimum value of k, see Fig. 2, which defines the particular point of operation, the yield p_1 can be calculated from its definition

$$p_1 = \frac{E_u}{E_t} \cdot \left[\frac{(T_u)_k}{(T_t)_k}\right]_{\text{operating point}} \tag{13}$$

To the extent that some abstracting journals may change their content within sections as well as their number of sections, from time to time, this analysis should be repeated at intervals over the period of publication of the abstracting journal.

4. EVALUATION OF THE YIELD P2

The establishment of a relation for the yield p_2 is not as simple as that for p_1. At least three different factors have to be taken into consideration:

1. Only a certain percentage of the references listed in the papers found through searching the abstracting journal for the first a years will be pertinent to the subject of interest. Let this percentage be denoted by J_I. In general one should expect $J_I \gg p_1$.
2. Some of the references listed in the bibliography of one particular paper will also be listed in others. The simplest way to take this duplication into account is to assume that it represents, on the average, a certain percentage J_{II} of the references of each paper located. Nevertheless, since finding of duplicate papers is an operation to be performed in any method of search, one need not include this item in the estimation of the cost of search by the present method. Hence, J_{II} is assumed to be equal to one.
3. A high percentage of the references listed in the bibliographies of papers located during the a years of search of the abstracting journal will be dated within the same period and consequently will already have been located. On the other hand, a certain percentage of these references will be dated more than b years back which means that probably they will be located again during the second and perhaps third cycle of the abstracting journal search. Estimation of this factor, which will be designated J_{III}, is obviously of importance.

Before an analysis of J_{III} is attempted some information is required on the distribution (by year) of all papers found in articles located through the abstracting journal search of *a* years.

For purposes of illustration consider that the papers published in 1953, on a specific subject, list a total of 464 references (excluding duplicate references). Their distribution with respect to date of publication is shown in Fig. 3. Such a graph can be normalized by dividing the ordinates by the total number of papers. If the total number of papers is large, and it were possible to order all papers by their date of publication, then the discrete points of Fig. 3 would give a continuous curve which can be anticipated to resemble that shown by solid line.

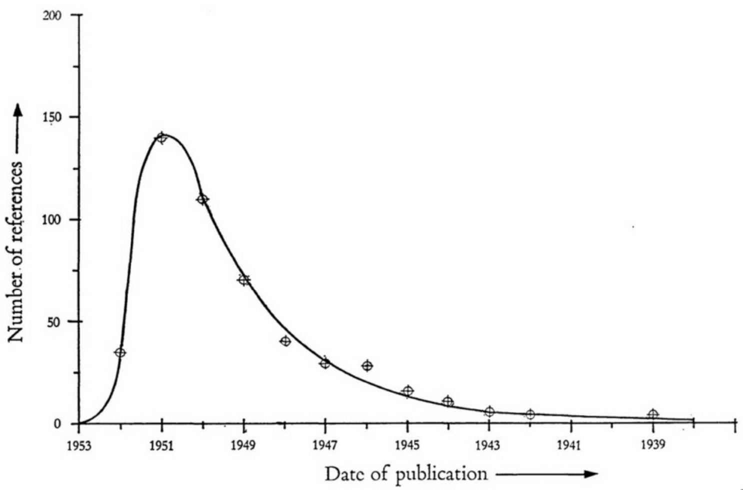


FIGURE 3. Distribution by date of useful references obtained from papers published in a given year, say 1953.

A preliminary study of the trend of a number of actual curves similar to the one of Fig. 3 has shown that the maximum occurs at about two years preceding the date of publication of papers whose references are listed.

In order to simplify the forthcoming analysis of J_{III} the curve of Fig. 3 is normalized and approximated by straight lines as shown in Fig. 4.

Let the interval *d* be divided into *z* equal parts. Then, if *P* is the ratio of the maximum ordinate in Fig. 3 divided by the total number of papers one can show that

$$\frac{P(1+z)}{2} = 1 \tag{14}$$

The parameters *c* and *d* have to be determined by searching through a number of papers pertinent to the subject studied. It will be seen later that the parameter, *P,* does not enter the calculations.

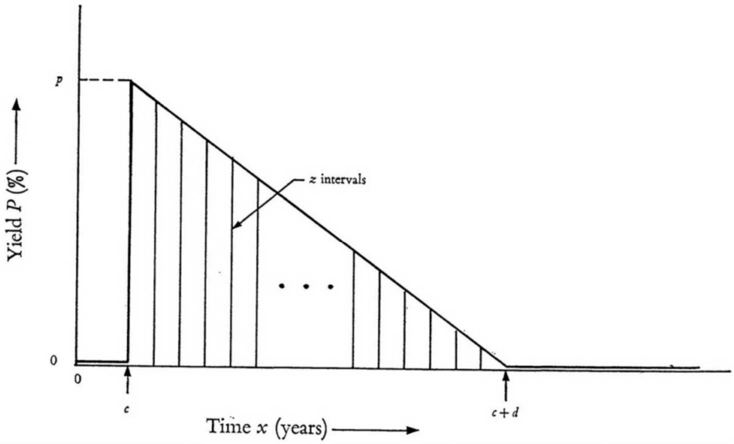


FIGURE 4. Simplified distribution model assumed in analysis.

The full lines of Fig. 5 show how the assumed distribution of Fig. 4 fits into the search cycle pattern of Fig. 1. As one can see, four different geometric positions are possible. The case for which $(c+d)<a$ is not shown since it is obvious that such a pattern is irrelevant to the analysis. The shaded areas of the triangles show the portion of references that are completely new. The point $x=0$, representing the commencement of abstracting journal search, gives the number of papers published in the same year as that of the abstracting journal. The point $x=a$ gives the number of papers published in the same year as that of the termination of the abstracting journal search.

Now let the interval *a* be divided into *z* equal subintervals *h,* so that

$$a=zh \tag{15}$$

Also let the shaded area of the triangle contained within the time interval $(b-a)$ be denoted by A_i. The ratio of the useful number of references to the total number of references obtained from the bibliographies is denoted by J_{III}. At this point in the analysis the assumption is made that the average number of papers published within the period of $(a+d)$ years is constant. With this

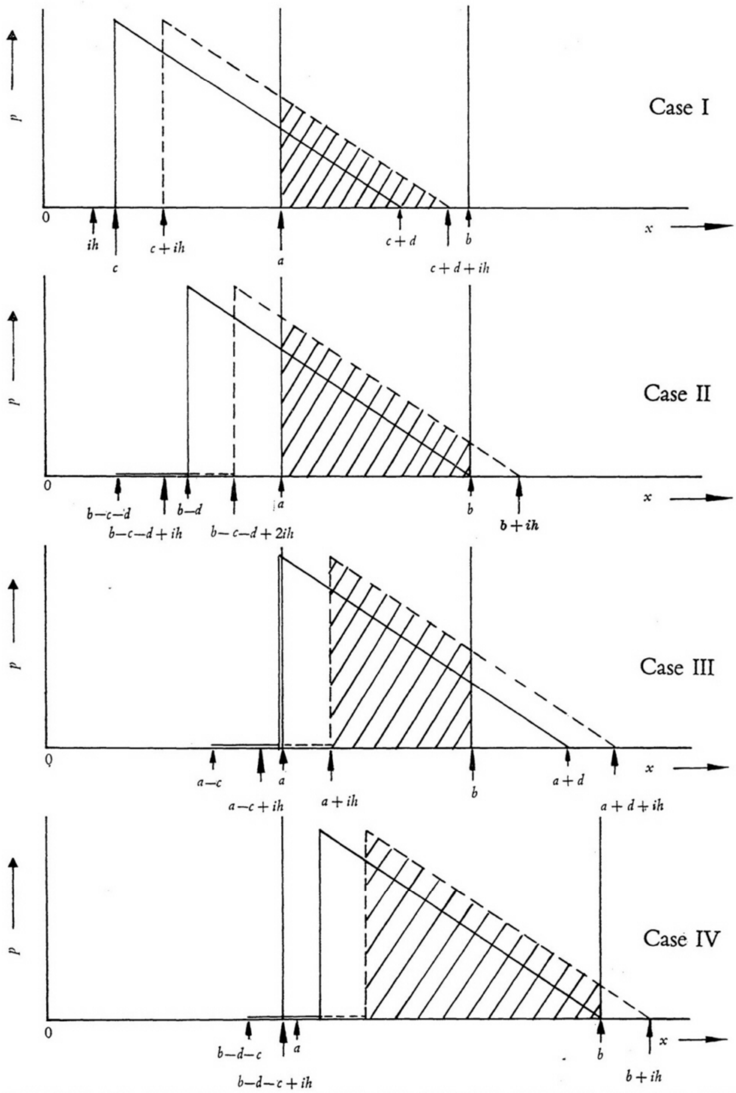


FIG. 5. Solid line triangles show overlap of references uncovered in cyclic search procedure based on parameters *a, b, c,* and *d*. Dotted line triangles indicate the same information. *(ih)* increment of time displaced.

understanding, which is true provided that *(a+d)* is not large, the following expression for J_{III} can be written:

$$J_{III} = \frac{\sum_{i=0}^{i=z} A_i}{z\left(\frac{Pd}{2}\right)} \qquad (16)$$

The detailed evaluation of J_{III} from Equation (16) is performed in Appendix A. This evaluation is assisted by a study of Fig. 5 where one observes that for increasing values of *x* the shapes of the shaded areas will be similar to cases I, II, and III if $d \geqq (b-a)$, or similar to cases I, IV, and III if $d \leqq (b-a)$. The first group, $d \geqq (b-a)$, is distinguished from the second, $d \leqq (b-a)$, by using primes on the pertinent parameters of the second group.

The calculations of Appendix A give J_{III} as a function of the ratios of the search parameters *a, b, c,* and *d,* for the two groups mentioned above. The results obtained for $d \geqq (b-a)$ are:

$$J_{III} = \frac{\left(\frac{b}{d}-\frac{c}{d}-1\right)^3 + 3\frac{a}{d}\left(\frac{b}{d}-\frac{a}{d}\right)\left(2+2\frac{c}{d}-\frac{b}{d}\right) - \left(\frac{c}{d}\right)^2\left(3+\frac{c}{d}\right)}{3\frac{a}{d}} \qquad (17)$$

and for $d \leqq (b-a)$ are:

$$J'_{III} = \frac{\left(\frac{a}{d}-\frac{c}{d}\right)^3 + 3\left(\frac{a}{d}-\frac{c}{d}\right)\left(\frac{c}{d}-\frac{a}{d}+1\right) + 3\frac{c}{d} - \left(\frac{a}{d}-\frac{b}{d}+\frac{c}{d}+1\right)^3}{3\frac{a}{d}} \qquad (18)$$

In conclusion, within the limitations of the conditions and assumptions imposed one can write by definition:

$$P_2 = J_I J_{III} \text{ for } d \geqq (b-a) \qquad (19)$$[4]

and

$$p`_2 = J_I J`_{III} \text{ for } d \leqq (b-a)$$

5. ADEQUACY OF PROPOSED SEARCH METHOD AND SELECTION OF PARAMETERS A AND B

As it was mentioned before, once the shape of the distribution curve of Fig. 3 is established, leading to the simplified Fig. 4, the parameters *c* and *d* are defined.

[4] Actually the yield, *p,* is given by the product of J_I, J_{II}, and J_{III}; however, earlier in the discussion J_{II} was assumed to be equal to one.

It is apparent from Equations (5) or (6) that small values of *a* result in small cost ratios, *R,* leading to a more economical search. On the other hand, for reasons that will become evident later, small values of *a* result in incomplete information during the interval *(b–a).* In general one can show that as the ratio *a/b* becomes progressively smaller, the ratio of the two costs, *R,* also progressively decreases.[5]

On the other hand, the extent to which *b* may increase is inherently limited, since, should *b* exceed *(a+c+d),* the interval *[b–(a+c+d)]* will not be covered by the references. Figure 5 shows that as *b* tends to equal *a* the time interval *(b–a)* receives better coverage by the references located in the bibliographies. Hence, a value of the ratio *a/b* has to be found which gives both a small cost ratio *R* and adequate coverage of the years *(b–a).*

In order to proceed with the determination of the appropriate value of *a/b,* the general distribution of the total collected information for the period *(b–a)* years has to be found. This determination is facilitated by reference to Fig. 6

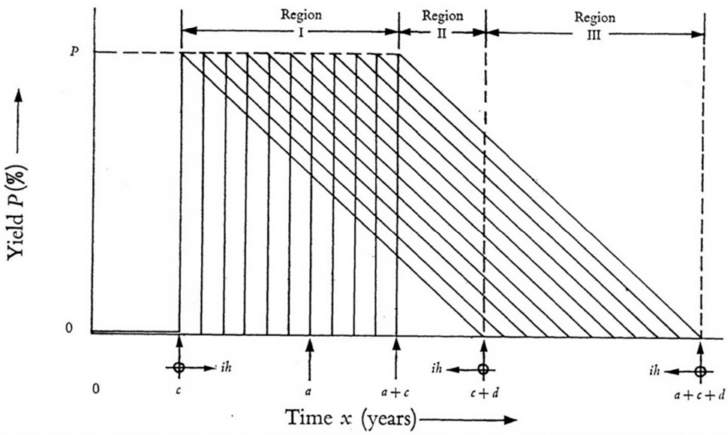


FIGURE 6. Geometric procedure of estimating distribution of useful references obtained from bibliographies of papers uncovered during *a* years of search in abstracting journal.

[5] To prove this statement differentiate Equation (5) with respect to *a/b*. This gives

$$\frac{dR_1}{d(a/b)} = \left(1 - \frac{p_1}{p_2}\right) + \left(1 - \frac{a}{b}\right)\frac{d(p_1/p_2)}{d(a/b)}.$$

In this equation, as *(a/b)* increases, p_2 decreases, since as *b* approaches *a,* greater portions of the extremities of the triangles of Fig. 5 are not used. Hence p_1/p_2 increases thus making $d(p_1/p_2)/d\text{-}(a/b)$ positive. Since p_1/p_2 and *a/b* are less than one, the derivative $dR_1/d(a/b)$ is always positive. (*Note.* In this footnote *d* stands for the differential sign and not as the search parameter.)

which essentially is an extension of the cases presented in Fig. 5. The regions I, II, and III of Fig. 6 are chosen arbitrarily to simplify the analytical summation processes involved by shifting the origins of the distribution curves. In this sense the limits of regions I, II, and III have no physical significance in the time, *x*, and their choice is only dictated by geometric considerations so that the mathematical analysis yields simple continuous functions within each region. The detailed steps in the summation of the ordinates of the triangles leading to the determination of the percentage of available information in a given year appear in Appendix B. The results of the calculations for the three regions are found to be:

$$S_{I,x} = 2\frac{(x-c)}{d}\left(1+\frac{c}{2d}-\frac{x}{2d}\right) \qquad c \leqq x \leqq a+c \tag{20}$$

$$S_{II,x} = \left(\frac{a}{d}\right)^2\left[1+2\frac{(d+c)}{a}-2\frac{x}{a}\right] \qquad c+a \leqq x \leqq c+d \tag{21}$$

$$S_{III,x} = \left(\frac{a+c+d}{d}-\frac{x}{d}\right)^2 \qquad c+d \leqq x \leqq c+d+a \tag{22}$$

For $x=a+c$, Equation (20) or (21) gives

$$S_{x=a+c} = 2\frac{a}{d}-\left(\frac{a}{d}\right)^2 = S_{max} \tag{23}$$

Equations (20) and (22) represent parabolas whereas Equation (21) represents a straight line. The quantity S_{max} is a number always smaller than one, approaching unity as a limit only if $a=d$. Figure 7 is a graphical representation of Equations (20) to (22) for the case $a=5$, $c=2$ and $d=7$, chosen arbitrarily but shown by experience to be representative values.

The parameters *a* and *b* can now be selected on the basis of the following argument. The maximum value of *S* will always be found at $x=a+c$, and will have a certain value, say $M\leqq 1$, whereas *b* will be chosen in such a manner that $S_{x=b}$ will not be smaller than a second value, say *N*, where again $N\geqq M\geqq 1$. For given values of *c, d,* and *M* the exact shape of the curve $S=S(x)$ in Fig. 7 is defined since from Equation (23)

$$M = 2\left(\frac{a}{d}\right)-\left(\frac{a}{d}\right)^2$$

hence,

$$a = d(1-\sqrt{1-M}) \tag{24}$$ [6]

[6] The second root of this equation falls at $x>(d+c)$. Instead of calculating *a* from the value of $S_{max}=S_{x=(a+c)}$, one can calculate *a* from the expression for *S* at $x=a$ given by Equation (20)

$$S_{x=a} = \frac{a(2d-a)}{(a-c)\,(2d-a+c)}$$

The value of *a* can be calculated from this equation by setting $S_x=a<1$; however, the resulting expression for *a/d* is more complicated than that obtained from Equation (24).

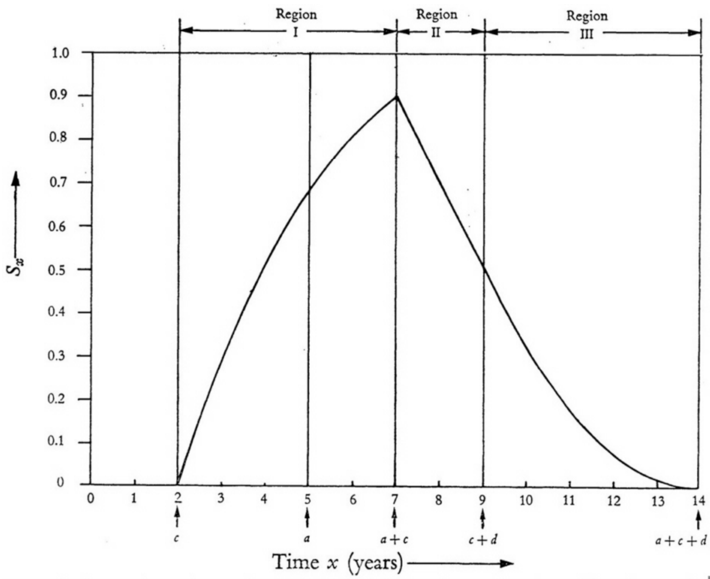


FIGURE 7. Result of summations of ordinates of Fig. 6 at each increment of time, *h,* for regions I, II, and III.

In order to calculate *b,* the value of *N* is selected so that $N=S_{x=b}<S_{x=a}$. As one can see from the equation in footnote 6, the value of $S_{x=a}$ is close to the maximum value $S_{x=(a+c)}$ since *c* in general is a small number (on the order of 2 years) compared with *a*.

The value of *x* for $S=N$ can fall either in region II or region III of Fig. 7. On the assumption that *N* lies in region II, $(c+a\leqq x=b\leqq c+d)$, one obtains from Equation (21)

$$N=\left(\frac{a}{d}\right)^2\left[1+2\frac{(d+c)}{a}-2\frac{x}{a}\right]$$

But, since $x=b$ by definition

$$b=x=\frac{a}{2}+(d+c)-\frac{Nd^2}{2a} \tag{25}$$

Substituting the value of *a* from Equation (24) in (25) gives

$$b=c+d\left[2-\frac{(M+N)}{2(1-\sqrt{1-M})}\right] \tag{26}$$

When b lies in region II, $b<(c+d)$, the relation between M and N as obtained from Equation (26) is

$$\frac{M+N}{1-\sqrt{1-M}} \geqq 2 \tag{27}$$

Now, for the case when b falls in region III Equation (22) gives

$$N = \left(\frac{a+d+c}{d} - \frac{x}{d}\right)^2.$$

Once again, since $x=b$ by definition

$$b = x = c + d(2 - \sqrt{1-M} - \sqrt{N}) \tag{28}$$

The new relationship between M and N for this case, as obtained from Equation (28) is

$$\frac{M+N}{1-\sqrt{1-M}} \leqq 2 \tag{29}$$

Having determined the values of a and b these can be substituted into the expression for J_{III} and the ratio R thus calculated.

6. SUMMARY OF RESULTS AND OPERATIONAL PROCEDURE

The results of the foregoing analysis are summarized below in a manner which illustrates the procedural steps followed in order to ascertain the appropriate search pattern to be used in an extensive survey of the literature in a given technical field.

1. Having clearly understood the nature and the scope of the subject under consideration, select the appropriate abstracting journal which gives a representative coverage of the world literature on this subject.

2. Starting with the most recent year of the abstracting journal selected, search through all the sections of the journal for the complete year.

3. Based on the useful abstracts uncovered in each section of the abstracting journal and the total number of abstracts in each section, calculate $(T_u)_k$ and $(T_t)_k$ from Equations (10) and (11) and construct the yield distribution curve as given in Fig. 2.[7]

$$(T_u)_k = \sum_{i=1}^{k} (Q_u)_i \qquad k = 1, 2, \cdots, s \tag{10}$$

$$(T_t)_k = \sum_{i=1}^{k} (Q_t)_i \qquad k = 1, 2, \cdots, s \tag{11}$$

[7] It is to be remembered that the k sections must first be ordered as specified by Equation (9).

4. From the information in step 3 determine the yield p_1 from Equation (13).

$$p_1 = \frac{E_u}{E_t}\left[\frac{(T_u)_k}{(T_t)_k}\right]_{\text{operating point desired}} \tag{13}$$

5. Obtain the papers represented by the useful abstracts uncovered in step 2 and search the bibliographies of all these papers for new references. From this information calculate J_I and construct Fig. 3.

6. Approximate the actual distribution curve of Fig. 3 by a triangle, thus obtaining the simplified model of Fig. 4. From Fig. 4 the search parameters *c, d,* and *c/d* are obtained.

7. Select the desired values of *M* and *N*.

8. With the selected values of *M* and *N* calculate the ratio $[(M+N)/(1-\sqrt{1-M})]$. If this quantity is greater than 2, then *b* falls in region II of Fig. 7. If this quantity is less than 2 then *b* falls in region III of Fig. 7.

For region II:
$$\frac{M+N}{1-\sqrt{1-M}} \geqq 2 \tag{27}$$

For region III:
$$\frac{M+N}{1-\sqrt{1-M}} \leqq 2 \tag{29}$$

9. Based on the information obtained from step 8 use Equation (26) or (28) to calculate the ratio *b/d*.

For region II:
$$\frac{b}{d} = \frac{c}{d} + \left[2 - \frac{M+N}{2(1-\sqrt{1-M})}\right] \tag{26}$$

For region III:
$$\frac{b}{d} = \frac{c}{d} + [2 - \sqrt{1-M} - \sqrt{N}] \tag{28}$$

10. By using Equation (24) calculate *a/d.*

$$\frac{a}{d} = 1 - \sqrt{1-M} \tag{24}$$

11. Calculate J_III or J'_III from Equations (17) or (18), depending on the sign of the inequality applicable to these equations. For $d \geqq b-a$:

$$J_\text{III} = \frac{\left(\frac{b}{d} - \frac{c}{d} - 1\right)^3 + 3\frac{a}{d}\left(\frac{b}{d} - \frac{a}{d}\right)\left(2 + 2\frac{c}{d} - \frac{b}{d}\right) - \left(\frac{c}{d}\right)^2\left(3 + \frac{c}{d}\right)}{3\frac{a}{d}} \tag{17}$$

For $d \leqq b-a$:

$$J'_{III} = \frac{\left(\frac{a}{d}-\frac{c}{d}\right)^3 + 3\left(\frac{a}{d}-\frac{c}{d}\right)\left(\frac{c}{d}-\frac{a}{d}+1\right) + 3\frac{c}{d} - \left(\frac{a}{d}-\frac{b}{d}+\frac{c}{d}+1\right)^3}{3\frac{a}{d}} \quad (18)$$

Also note that the inequality $d \leqq b-a$ can be transformed into the more convenient form $[1-(b/d)+(a/d)]>0$.

12. With J_I and J_{III}, or J'_{III}, as the case may be, calculate p_2 or p'_2 from Equation (19).

$$P_2 = J_I \cdot J_{III} \text{ for } d \geqq b-a$$
$$p'_2 = J_I \cdot J'_{III} \text{ for } d \geqq b-a \quad (19)$$

13. Finally, using Equations (5) or (6), calculate R_1 or R_2 depending upon the type of search cycle used.

For integral number of cycles:

$$R_1 = \frac{a}{b} + \left(1-\frac{a}{b}\right)\frac{p_1}{p_2} \quad (5)$$

For non-integral number of cycles:

$$R_2 = \frac{\left(1+\frac{1}{n_2}\right)\cdot\left[\frac{a}{(b-a)}\right] + \frac{p_1}{p_2}}{\left(1+\frac{1}{n_2}\right)\cdot\left[\frac{a}{(b-a)}\right] + 1} \quad (6)$$

7. ILLUSTRATIVE EXAMPLE

The following factual example serves to illustrate the advantages to be derived through the use of the search procedure developed in the present analysis.

SUBJECT OF INTEREST

It is proposed to conduct an extensive literature search of the thermophysical properties of metals. The thermal conductivity, specific heat, thermal emissivity and the thermal diffusivity of both ferrous and non-ferrous metals and their alloys are of interest.

INITIAL CONSIDERATIONS

The type of information of interest is covered, rather extensively, in at least three well-known abstracting journals, namely *Chemical Abstracts, ASM Review of Metals Literature,* and *Metallurgical Abstracts*. Suppose *Metallurgical Abstracts* is selected as the source journal to be used in the search of the literature. The question to be answered is: "What is the optimum method for searching the complete series of the *Metallurgical Abstracts* starting with the year 1957 and going all the way back to 1934, its first year of publication."

PROCEDURE OF ANALYSIS

Following the step by step procedure outlined in Sec. 6 the following quantities are determined:

1. An item by item search of Volume 22 (September 1954 through August 1955) of *Metallurgical Abstracts,* consisting of 28 sections gives the approximate curve *A* of Fig. 2.

2. The search of this volume yields the quantities E_u=262 abstracts and E_t=7850 abstracts.

3. From the search it is established that $(T_u)_k$=0.87 and $(T_t)_k$=0.29.

4. By using Equation (13):[8]

$$p_1 = \frac{E_u}{E_t}\cdot\frac{(T_u)_k}{(T_t)_k} = \frac{262}{7850}\cdot\frac{0.87}{0.29} = 0.10$$

5. From the references listed in a limited number of papers relevant to the subject under investigation it is estimated that J_I=0.80.

6. From the same investigation as in step 5, it is estimated that c/d=0.2.

7. It is planned to conduct the literature search for values of M=0.90 and N=0.40.

8. Calculate the quantity

$$\frac{M+N}{1-\sqrt{1-M}} = \frac{0.90+0.40}{1-\sqrt{1-0.90}} = 1.9 < 2$$

9. Since, from step 8, 1.9<2, Equation (28) is used to calculate b/d.

$$\frac{b}{d} = \frac{c}{d} + (2-\sqrt{1-M}-\sqrt{N}) = 0.2 + (2-\sqrt{1-0.9}-\sqrt{0.4}) = 1.25$$

10. From Equation (24):

$$\frac{a}{d} = 1-\sqrt{1-M} = 1-\sqrt{1-0.9} = 0.683$$

11. Prior to the calculation J_{III} or J'_{III} investigate the sign of the quantity

$$(1-\frac{b}{d}+\frac{a}{d}) = 1-1.25+0.683 = 0.433 > 0$$

hence,

$$d>(b-a)$$

---

[8] The operating point for this volume was chosen at k=8. The specific sections selected for further search were: 1, 2, 4, 5, 12, 26, 27, and 28.

One concludes from this that Equation (17) is to be used to calculate J_{III}.

$$J_{III} = \frac{\left(\frac{b}{d} - 1 - \frac{c}{d}\right)^3 + 3\frac{a}{d}\left(\frac{b}{d} - \frac{a}{d}\right)\left(2 + 2\frac{c}{d} - \frac{b}{d}\right) - \left(\frac{c}{d}\right)^2\left(3 + \frac{c}{d}\right)}{3\frac{a}{d}}$$

$$J_{III} = \frac{(1.25-1-0.2)^3 + 3(0.683)\,(1.25-0.683)\,(2 + 2 \times 0.2-1.25)}{3(0.683)} - \frac{(0.2)^2(3 + 0.2)}{3(0.683)} = 0.59$$

12. $p_2 = J_I J_{III} = 0.80 \times 0.59 = 0.47$
13. For a search procedure using an integral number of cycles

$$R_1 = \frac{a}{b} + \left(1 - \frac{a}{b}\right)\frac{p_1}{p_2}$$

or

$$R_1 = \frac{a/d}{b/d} + \left(1 - \frac{a/d}{b/d}\right)\frac{p_1}{p_2}$$

$$= \frac{0.683}{1.25} + \left(1 - \frac{0.683}{1.25}\right)\frac{0.10}{0.47} = 0.65$$

This final result indicates that, for the case investigated, the cost of the cyclic method of search is only 65% of the cost of the conventional way of searching by using the abstracting journal only. In other words, a 35% reduction in search cost is realized.

In concluding this illustrative example it may be worth noting that for the case under consideration the abstracting journal search and bibliographic search periods are taken nearly equal (a/b=0.546). Based on the observation in the text that a should not be greater than 5 years, and considering that Y=23 years in this example, Equation (1) gives bn_1=23. Since a, b, and n_1 are preferably whole numbers, one can select for the search cycle the combination:

$$a=3,\ b=6,\ n_1=4$$

or

$$a=2,\ b=4,\ n_1=6$$

If the first set of values are used, the *Metallurgical Abstracts* will be searched in four cycles with a=3 years and $(b-a)$=3 years.

APPENDIX A DETERMINATION OF JIII IN TERMS OF SEARCH PARAMETERS A, B, C AND D

The following analysis refers to Fig. 5 of Sec. 4 leading to the determination of Equations (17) and (18). In this figure let z_1, z_2, z_3 and $z_1`$, $z_2`$, $z_3`$ represent the number of time intervals h for which the areas A_i correspond to cases I, II, III and I, IV, III, respectively. With this notation one can write:

and
$$z_1 + z_2 + z_3 = z'_1 + z'_2 + z'_3 = z = \frac{a}{h} \tag{A-1}$$

$$\left.\begin{aligned} hz_1 &= (b - d - c) \\ hz_2 &= (a - b + d) \\ hz_3 &= c \end{aligned}\right\} \quad d \geqq (b - a) \tag{A-2}$$

$$\left.\begin{aligned} hz'_1 &= (a - c) \\ hz'_2 &= (b - a - d) \\ hz'_3 &= (a + c + d - b) \end{aligned}\right\} \quad d \leqq (b - a) \tag{A-3}$$

For the case of $d \geqq (b-a)$, Equation (A-2), and by referring to Fig. 5, one can write by inspection pertinent relations for each of the three cases.

For case I:

$$A_i = \frac{P}{2d}(c + d - a + ih)^2 = \frac{P}{2d}\left[(c + d - a)^2 + 2\,(c + d - a)\,ih + (ih)^2\right]$$

Hence,

$$\frac{\sum_{i=1}^{z_1} A_i}{z(Pd)/2} = \frac{1}{zd^2}\,[(c + d - a)^2\,z_1 + 2\,(c + d - a)\,(1 + 2 +, \cdot \cdot \cdot, + z_1)\,h + (1^2 + 2^2 +, \cdot \cdot \cdot, + z_1^2)h^2]$$

$$= \frac{1}{d^2}\left(\frac{z_1}{z}\right)\left[(c + d - a)^2 + a\,(c + d - a)\left(\frac{z_1}{z} + \frac{1}{z}\right) + \frac{a^2}{6}\left(2\frac{z_1}{z} + \frac{1}{z}\right)\left(\frac{z_1}{z} + \frac{1}{z}\right)\right] \tag{A-4}$$

For case II:

$$A_i = \frac{P}{2d}\,(b - a)\,(b - a + 2ih)$$

Hence,

$$\frac{\sum_{i=1}^{z_2} A_i}{z(Pd)/2} = \frac{1}{zd^2}\,(b - a)\,[(b - a)\,z_2 + 2h\,(1 + 2 +, \cdot \cdot \cdot, + z_2)]$$

$$= \frac{(b - a)}{d^2}\left(\frac{z_2}{z}\right)\left[(b - a) + a\left(\frac{z_2}{z} + \frac{1}{z}\right)\right] \tag{A-5}$$

For case III:

$$A_i = \frac{P}{2d}[(2d+a-b)+ih][(b-a)-ih] = \frac{P}{2d}[(2d+a-b)(b-a)-2(a+d-b)ih-(ih)^2]$$

Hence,

$$\frac{\sum_{i=1}^{z_3} A_i}{z(Pd)/2} = \frac{1}{zd^2}[(2d+a-b)(b-a)z_3 - 2(d+a-b)h(1+2+,\cdots,+z_3) - h^2(1^2+2^2+,\cdots,+z_3^2)]$$

$$= \frac{1}{d_2}\left(\frac{z_3}{z}\right)\left[(2d+a-b)(b-a) - a(d+a-b)\left(\frac{z_3}{z}+\frac{1}{z}\right) - \frac{a^2}{6}\left(2\frac{z_3}{z}+\frac{1}{z}\right)\left(\frac{z_3}{z}+\frac{1}{z}\right)\right] \quad \text{(A-6)}$$

Now the ratio J_{III} is given by the sum of Equations (A-4), (A-5), and (A-6). In these expressions the time interval, *h,* can be arbitrarily chosen to be very small, making *z* a large number. Consequently the terms $1/z$ appearing in these equations can be neglected. With this simplification and by substituting for the ratios z_1/z, z_2/z, z_3/z from Equation (A-2) and by using Equation (14), one obtains for $d \geqq b-a$

$$J_{III} = \frac{(b-d-c)^3 + 3a(b-a)(2d+2c-b) - c^2(3d+c)}{3ad^2} \quad \text{(A-7)}$$

Dividing Equation (A-7), top and bottom, by $(d)^3$ one obtains

$$J_{III} = \frac{\left(\frac{b}{d}-\frac{c}{d}-1\right)^3 + 3\frac{a}{d}\left(\frac{b}{d}-\frac{a}{d}\right)\left(2+2\frac{c}{d}-\frac{b}{d}\right) - \left(\frac{c}{d}\right)^2\left(3+\frac{c}{d}\right)}{3(a/d)} \quad \text{(A-8)}$$

Equation (A-8) is the result given as Equation (17) in the text.

The expression for J_{III}, when $d \leqq (b-a)$, Equation (A-3), can be determined in a manner similar to that above. However, in this instance one can utilize some of the results already obtained.

The relation corresponding to Equation (A-4) can be obtained by replacing z_1 by z'_1 in this equation, resulting in

$$\frac{\sum_{i=1}^{z'_1} A_i}{z(Pd)/2} = \frac{1}{d^2}\left(\frac{z'_1}{z}\right)[(c+d-a)^2 + a(c+d-a)\left(\frac{z'_1}{z}+\frac{1}{z}\right) + \frac{a^2}{6}\left(2\frac{z'_1}{z}+\frac{1}{z}\right)\left(\frac{z'_1}{z}+\frac{1}{z}\right) \quad \text{(A-9)}$$

The relation corresponding to Equation (A-5) follows from an examination of Fig. 5, case IV, from which it can be established that

$$\frac{\sum_{i=1}^{z'_2} A_i}{z(Pd)/2} = \frac{z'_2}{z} \quad \text{(A-10)}$$

The equation corresponding to Equation (A-6) is obtained by again examining Fig. 5, case IV, from which one can write

$$A_i = \frac{P}{2d}(d - ih)(d + ih) = \frac{P}{2d}[d^2 - (ih)^2]$$

Hence,

$$\frac{\sum_{i=1}^{z'_3} A_i}{z(Pd)/2} = \frac{1}{zd^2}[d^2 z'_3 - h^2 (1 + 2^2 +, \cdots, + z'^2_3)]$$

$$= \frac{1}{d^2}\left(\frac{z'_3}{z}\right)\left[d^2 - \frac{a^2}{6}\left(2\frac{z'_3}{z} + \frac{1}{z}\right)\left(\frac{z'_3}{z} + \frac{1}{z}\right)\right] \quad \text{(A-11)}$$

Again, by summing up equations (A-9), (A-10), and (A-11), neglecting $1/z$, and substituting for the quantities z'_1/z, z'_2/z, z'_3/z from Equation (A-3) one obtains for $d \leqq b-a$

$$J'_{\mathrm{III}} = \frac{(a-c)^3}{3ad^2} + \frac{(a-c)(c+d-a)}{ad} + \frac{c}{a} - \frac{(a+c+d-b)^3}{3ad^2} \quad \text{(A-12)}$$

$$J'_{\mathrm{III}} = \frac{\left(\frac{a}{d} - \frac{c}{d}\right)^3 + 3\left(\frac{a}{d} - \frac{c}{d}\right)\left(\frac{c}{d} - \frac{a}{d} + 1\right) + 3\frac{c}{d} - \left(\frac{a}{d} - \frac{b}{d} + \frac{c}{d} + 1\right)^3}{3\,(a/d)} \quad \text{(A-13)}$$

Equation (A-13) is the result given as Equation (18) in the text.

APPENDIX B EVALUATION OF SX AS A FUNCTION OF SEARCH PARAMETERS A, B, C, D AND TIME, X

In order to proceed with the calculation of S_x one must refer to Fig. 6 and consider the regions I, II, and III depicted therein. To facilitate the summation procedure within each of these regions, it is found desirable to locate the origin for *x* at the points *c, (c+d),* and *(a+c+d),* respectively. The direction of increasing *x* for each region is indicated by arrows in Fig. 6.

Following the practice as outlined in the text (see Equation 15), let the time interval *a* be divided into *z* equal parts of magnitude *h*. The summation of all ordinates of the triangles at each specific time corresponding to a specific value off within the regions I, II, and III is represented by $S_{\mathrm{I},i}$, $S_{\mathrm{II},i}$, and $S_{\mathrm{III},i}$, respectively.

Summations for region I are:

$$S_{\mathrm{I},0} = P$$

$$S_{\mathrm{I},1} = P + \left(P - \frac{Pa}{dz}\right) = P\left(2 - \frac{a}{dz}\right)$$

$$S_{\mathrm{I},2} = S_{\mathrm{I},1} + \left(P - 2\frac{Pa}{dz}\right) = P\left(3 - \frac{a\,(1+2)}{dz}\right)$$

. . . .

$$S_{\mathrm{I},i} = S_{\mathrm{I},(i-1)} + \left(P - i\frac{Pa}{dz}\right) = P\left[(i+1) - \frac{(1 + 2 +, \cdots, + i)\,a}{dz}\right]$$

$$= P(1+i)\left(1-\frac{ia}{2dz}\right) \tag{B-1}$$

. . . .

$$S_{I,z} = P(1+z)\left(1-\frac{a}{2d}\right) \tag{B-2}$$

Summations for region III are:

$$S_{III,0} = 0$$

$$S_{III,1} = \frac{aP}{dz}$$

$$S_{III,2} = \frac{aP}{dz}(1+2)$$

. . . .

$$S_{III,i} = \frac{aP}{dz}(1+,\cdots,+i) = \frac{i(i+1)}{2}\frac{aP}{dz} \tag{B-3}$$

. . . .

$$S_{III,z} = \frac{aP}{2d}(1+z) \tag{B-4}$$

Summations for region II are:

$$S_{II,0} = S_{III,z} = \frac{aP}{2d}(1+z)$$

$$S_{II,1} = S_{III,(z+1)} - S_{III,0} = \left[\frac{(z+1)(z+2)}{2} - 0\right]\frac{aP}{dz}$$

. . . .

$$S_{II,i} = S_{III,z+i} - S_{III,(i-1)} = \left[\frac{(z+i)(z+i+1)}{2} - \frac{i(i-1)}{2}\right]\frac{aP}{dz}$$

$$= \frac{(z+1)(z+2i)}{2}\cdot\frac{aP}{dz} \tag{B-5}$$

. . . .

$$S_{II,z[(d/a)-1]} = \frac{(1+z)(2d-a)P}{2d} \tag{B-6}$$

The range of variation of the indices *h* in the three regions is as follows:

Region I:	$0<i/z<1$
Region II:	$0<i/z<(d/a)-1$
Region III:	$0<i/z<1$

By inspection of Fig. 6, one observes that the magnitude of $S_{II,z}$ $[_{(d/a)-1}]$, corresponding to the point $x=(a+c)$, is the maximum of all S_x values in the three regions.

The quantities S_i are by their definition proportional to the number of references redeemed in each specific year by searching the bibliographies of the papers located from the search of the abstracting journal. The maximum number of references that could have been compiled by this process is proportional to the summation of all ordinates of the assumed distribution triangle. As it was found previously in Equation (14),

$$\frac{P}{2}\left(1+\frac{dz}{a}\right)=1 \tag{B-7}$$

Now, dividing Equations (B-1), (B-3), and (B-5) by Equation (B-7) gives

$$S_{\mathrm{I},i} = \left(\frac{S_{\mathrm{I},i}}{\frac{P}{2}\left(1+\frac{dz}{a}\right)}\right)_{z\to\infty} = \frac{P\,(1+i)\left(1-\frac{ia}{2zd}\right)}{\frac{P}{2}\left(1+\frac{dz}{a}\right)} = 2\cdot\frac{a}{d}\cdot\frac{i}{z}\left(1-\frac{1}{2}\cdot\frac{i}{z}\cdot\frac{a}{d}\right) \tag{B-8}$$

$$S_{\mathrm{II},i} = \left(\frac{S_{\mathrm{II},i}}{\frac{P}{2}\left(1+\frac{dz}{a}\right)}\right)_{z\to\infty} = \frac{\frac{aP}{2dz}(1+z)\,(z+2i)}{\frac{P}{2}\left(1+\frac{dz}{a}\right)} = \left(\frac{a}{d}\right)^2\left(1+2\frac{i}{z}\right) \tag{B-9}$$

$$S_{\mathrm{III},i} = \left(\frac{S_{\mathrm{III},i}}{\frac{P}{2}\left(1+\frac{dz}{a}\right)}\right)_{z\to\infty} = \frac{\frac{aP}{2dz}i\,(i+1)}{\frac{P}{2}\left(1+\frac{dz}{a}\right)} = \left(\frac{a}{d}\right)^2\left(\frac{i}{z}\right)^2 \tag{B-10}$$

At this point in the development it becomes once again convenient to adopt a common origin, x=0, using the distance, x, from the origin for all three regions instead of different values of i/z. By performing the appropriate transformation one obtains

For Region I:

$$\frac{i}{z}=\frac{x-c}{a}$$

For Region II:

$$\frac{i}{z}=\frac{d+c-x}{a}$$

For Region III:

$$\frac{i}{z}=\frac{a+d+c-x}{a}$$

Using these substitutions in Equations (B-8), (B-9) and (B-10) gives the new set of equations:

$$S_{\mathrm{I},x} = 2\frac{(x-c)}{d}\left(1+\frac{c}{2d}-\frac{x}{2d}\right) \qquad c \leqq x \leqq a+c \tag{B-11}$$

$$S_{\mathrm{II},x} = \left(\frac{a}{d}\right)^2\left[1+2\frac{(d+c)}{a}-2\frac{x}{a}\right] \qquad c+a \leqq x \leqq c+d \tag{B-12}$$

$$S_{\mathrm{III},x} = \left(\frac{a+c+d}{d}-\frac{x}{d}\right)^2 \qquad c+d \leqq x \leqq c+d+a \tag{B-13}$$

Equations (B-11), (B-12), and (B-13) are the results given as Equations (20), (21), and (22) in the text.

NOMENCLATURE

Symbol	*Definition*
a	Number of years of literature search through abstracting journal.
A_i	Area of triangles in Fig. 6, bounded by search lengths of a and b years.
b	Number of years covered by a complete search cycle.
c, d	Search parameters, points (years) in search period.
C	Cost of searching through n number of repetitive cycles as shown in Fig. 1.
C_a	Cost of searching by going through abstracting journal only.
E	Number of abstracts in abstracting journal or bibliography at end of a paper.
h	Increment of time within period of *(b`a)* years.
i	A positive integer.
J_I	Yield of useful references obtained from the bibliography of a paper originally located from abstracting journal.
J_{II}	Average per cent of references obtained from the bibliography of a paper which are found duplicated in the bibliographies of other papers.
J_{III}	Ratio of useful references to the total number of references located in the bibliographies of all papers found during search of abstracting journal for a period of a years.
K	Proportionality factor.
M, N	Particular values of the summations S_x.
n	Number of search cycles.
P_1	Per cent of useful abstract yield in searching through abstracting journal.
P_2	Per cent of useful abstract yield in searching through bibliographies of technical papers.
P	Ratio of maximum ordinate of Fig. 3 divided by total number of papers.
Q	Ratio of abstracts defined by Equations (7) and (8).
R	Ratio of actual cost by cyclic search to cost of search by going through abstracting journal only, $R=C/C_a$.
s	Total number of sections in a given abstracting journal.
S	Summation of all ordinates of Fig. 6 at a given time, x.
T	Partial sum defined by Equations (10) and (11).
x	Time, in years.
Y	Length of total search period, in years.
z	Defined as a/h.
Subscripts	
$(\ )_1$	Refers to method of search using an integral number of cycles.
$(\ )_2$	Refers to method of search where search both commences and terminates using the abstracting journal (last cycle incomplete).
$(\ )_{i,k}$	Refers to the ith or kth item of a series of consecutive numbers.
$(\ )_t$	Refers to total number of abstracts.
$(\ )_u$	Refers to useful number of abstracts.
Superscripts	
( )`	Indicates the value of a parameter for the case $d \geqq (b-\bar{a})$ as specified in Appendix A.

The Relation Between Completeness and Effectiveness of a Subject Catalogue

C.S.SABEL

Anyone who has frequently undertaken literature searches will have used the references in material already found as a lead to further material relevant to the subject in which the search is being made. This experience suggests that it might be interesting to investigate the material that can be retrieved in this manner in order to see whether too much effort was perhaps being put into aiming at 100% storage of material and at 100% retrieval of the items (regarded as documents) stored.

The preliminary investigation described here analysed the references contained in the documents from three different sources dealing with the same subject (Controlled thermonuclear reactions).

The A.E.R.E. unclassified reports on radioactivation analysis were also examined to see how far the results from these reports agreed with those obtained from the documents on controlled thermonuclear reactions.

ANALYSIS OF REFERENCES

The 98 documents on controlled thermonuclear reactions studied in this investigation were in three classes. These were: *(a)* 27 Atomic Energy Research Establishment unclassified reports, (b) 20 published articles by A.E.R.E. authors, other than unclassified reports, *(c)* 51 United States Atomic Energy Commission unclassified reports, listed in a bibliography prepared by the U.S.A.E.C.

For the purpose of this study, these could be regarded as representing, in each class, a 100% sample of the documents dealing with the subject.

The number of times a document was quoted as a reference in other documents is set out in Table 1, which shows, for example, seventy-seven documents were not quoted as a reference in any other document, twelve documents were quoted as a reference in one other document, etc.

C.S.SABEL Atomic Energy Research Establishment, Harwell, England.

TABLE 1 All documents on controlled thermonuclear reactions

No. of documents	*Times quoted as reference*
77	0
12	1
3	2
6	3

Considering separately within themselves the three classes of documents on controlled thermonuclear reactions, we have Tables 2–4.

TABLE 2 A.E.R.E. unclassified reports

No. of documents	*Times quoted as reference*
19	0
4	1
4	2

TABLE 3 A.E.R.E. author's papers

No. of documents	*Times quoted as reference*
11	0
4	1
5	2

TABLE 4 U.S.A.E.C. unclassified reports

No. of documents	*Times quoted as reference*
49	0
2	1

The references in A.E.R.E. unclassified reports on radioactivation analysis, studied as a comparison, gave the breakdown shown in Table 5.

TABLE 5

No. of documents	*Times quoted as reference*
19	0
4	1
0	2
0	3
1	4

A 25th report, a bibliography, was excluded from Table 5 as the presence of a bibliography in a subject field will obviously have a considerable influence on a search, provided its existence is known. In this case, as the bibliography was recent, none of the other reports included it as a reference. The bibliography included 17 of the 24 reports.

CONSIDERATION OF TABLES 1 TO 5

The very high proportion of documents which are not included as references in any other document suggests that a complete retrieval from only some of the documents in a given category is unlikely.

The results were a disappointment from the statistical viewpoint in that it was hoped that they would be amenable to an analysis as, possibly, a binomial distribution. In which case $p=k/n$ where p is the probability of one document being quoted in another document, n is the total number of documents, and k is a constant. With a value for k, it would be possible by weighting the above tables with results from tables of the frequency of quotations in documents to arrive at, say, a 95% certainty of obtaining 100% of the documents by choosing some number less than the total number of documents and examining the references within these documents. However, the sample above was not large enough for one to be statistically categorical and the values of k that were obtained were not consistent. Intuitively it does appear to be true that the probability of a document being quoted in another document is inversely proportional to the total number of documents.

FURTHER ANALYSIS

In addition, the results were studied to see whether chains of references existed leading from a few to many documents, and to see how far references in one class of document led to documents in another class.

It was found that for the documents on controlled thermonuclear reactions there were no “key” references which led to a large number of others in the subject, and no document was quoted as a reference in each of the three classes of document. In particular, references tended to be restricted to their own class of document. For example, there was no reference in the A.E.R.E. reports or published papers to give entry to the U.S.A.E.C. reports and only one reference in the U.S.A.E.C. reports to an A.E.R.E. published paper.

CONCLUSIONS

Tables 1 to 5 show that the references in an incomplete list of documents are unlikely to indicate more than a small proportion of the remaining documents, and hence one is unlikely to be justified in retrieving less than all the documents that can be found from a subject catalogue.

The further analysis of the references also shows that, in the subject studied, there is no possibility of selecting documents for storage which would in themselves indicate most of the material one would wish to retrieve. The results also indicate the necessity, in planning an information service, of ensuring adequate retrieval of all types of relevant documents.

This is only a preliminary survey and further subject fields and different types of documents should be analysed to see if the restriction of references to documents within their own class is as marked in other fields.

Cost Analysis of Bibliographies or Bibliographic Services

MALCOLM RIGBY and MARIAN K.RIGBY

ABSTRACT. This report represents the tentative results of preliminary work done under a grant from the National Science Foundation for the purpose of establishing under controlled conditions an empirical formula for obtaining estimates of the order of magnitude of the costs involved in preparation of either "one-shot" bibliographies or more extensive or continuing services, regardless of quantity, accuracy, exhaustiveness, subject matter, languages involved, country where prepared, type or frequency of indexes, quality of annotations or reproduction, etc.

As the material produced in this analysis is quite voluminous and the exact figures for a number of the examples are not yet obtained, the tables and the curves on the nomogram should be considered as rough approximations by anyone not wishing to examine the "exhibits" (which will be available to those desiring to verify the cost or the quality of the product). Costs may vary by 200 or 300% from the "normal," on account of unpredictable factors, but in general they range between 50c and $50 per item in the extreme, and between $1 and $15 per item under "average" conditions.

The explanation of the range in costs of from 1 to 2 orders of magnitude lies not only in the complexity, exhaustiveness, or sophistication involved, but to a large measure in the "entropy" of the system; i.e., in the *amount and skill of the work previously done* by some library or bibliographer and available for exploitation by others; or, conversely on the amount of effort which is expended by the farsighted in putting material in shape for others to exploit to their advantage or to the advantage of science.

Finally one of the most significant factors, which usually operates in direct opposition to that which would be predicted, is the rising cost with increasing volume or size of a bibliography, for the extra controls, tools, possible mechanization, research, development of codes or class systems for large projects more than offsets any savings due to "streamlining."

MALCOLM RIGBY and MARIAN K.RIGBY American Meteorological Society, Washington, D.C.

PURPOSE

The object of this study is to obtain some preliminary empirical data on the costs of various types of bibliographies or bibliographic services. The final product, when supplemented and refined, will be a nomogram or a series of curves giving the approximate cost of a bibliography of any given size or degree of sophistication. The present study will merely provide a rough and general idea of the factors or variables involved, and the relative order of magnitude of the resulting costs. Finally it will permit an analysis of the proportion of the cost of each type of bibliography that would be chargeable to any given step in its preparation.

FACTORS INVOLVED IN VARIABLE COSTS

It is easily recognized by any experienced research worker that bibliographies vary as much in quality and usefulness as do any other professional services. What is not always recognized is the magnitude of the variation in unit cost as the exhaustiveness, language coverage, type of annotation, comprehensiveness of indexing or even size (volume) of the project increases. No one would question the fact that accuracy of entries, type of reproduction or country of preparation would make a considerable difference in unit cost, but the actual amount of this difference would be sheer guesswork without some quantitative data based on an actual comparison involving duplication of work at home and abroad.

Method of approach

It seemed advisable to eliminate as many as possible of the intangible variables (e.g., experience, diligence, availability of material, supervision, overhead, and unpredicted demands for irrelevant services which greatly increase the cost of a small project). This was done by preparing a series of bibliographies of approximate “unit” size, under uniform and controlled supervision, and with personnel of average and known experience and performance, and with uniform conditions and facilities for work. A unit was established as 100 items, and reproduction was set at 100 copies to eliminate the factor of reduction in cost with increasing number of copies printed. In some cases 200 or more items were included to make the work comparable in exhaustiveness with another similar work, but the basic unit cost would not change much between 100 and 200 items.

Bibliographies on the same subject, "Noctilucent clouds," were prepared by expert bibliographers simultaneously in the United States, England, Belgium, and Japan, to see what difference in cost would be involved and what difference in quality, if any, would be realized. The results can be seen in the "exhibits" by anyone interested.

Finally the analysis of various factors entering into the costs and other pertinent data were made possible by developing an analysis form (see Appendix 1) which would take into account the portion of the total cost which could be assigned to each step in the process of preparation of the bibliography. In some cases one or more of the processes (organizing, searching, cataloging, arranging, classifying, indexing, annotating or translating abstracts, editing, typing and proofing, duplicating and printing, overhead, etc.) would be missing or not pertinent, in which case the space is left blank. The results of the analysis are shown on Table 1.

TABLE 1 Unit cost of various types of bibliography of a given size

Type[a]	*Characteristic*	*10²* *(100)*	*10³* *(1000)*	*10⁴* *(10,000)*	*10⁵* *(100,000)*	*10⁶* *(1,000,000)*
I	Simple copying	$0.55	$0.60	$0.70	$0.60	$0.60
II	Arrange and copy	.85	1.30	1.80	1.50	1.30
III	Search and arrange	1.80	2.50	3.60	2.60	2.00
IV	Index in addition	2.30	4.00	5.00	4.00	3.00
V	Abstract and index	8.00	10.00	15.00	13.00	10.00
VI	Translate abstracts	5.00	9.00	12.00	18.00	20.00
VII	Mechanize abstracts	50.00	25.00	22.00	20.00	15.00
VIII	Reference and bibliographic	10.00	15.00	18.00	15.00	12.00

[a] Types I to IV are unannotated; V to VIII are annotated (abstracted).

It is hoped that this form can be used to accumulate further data on a variety of projects not yet studied. In some cases the figures may be considered confidential, yet the "unit cost" could be entered in the revised cost analysis table and could affect the curves on the nomogram, thus making them more meaningful.

TYPES OF BIBLIOGRAPHIES

Figure 1 shows eight curves representing the eight basic types of bibliographies designated below and indicating the approximate cost per item for the range from 10^2 to 10^6 items (or items per year), under conditions considered normal in the United States. Costs for the first eight types are based on results of work done on this project, projected from known costs for the basic unit of 100 items to 1000, 10,000 or more items from approximate costs of known projects on

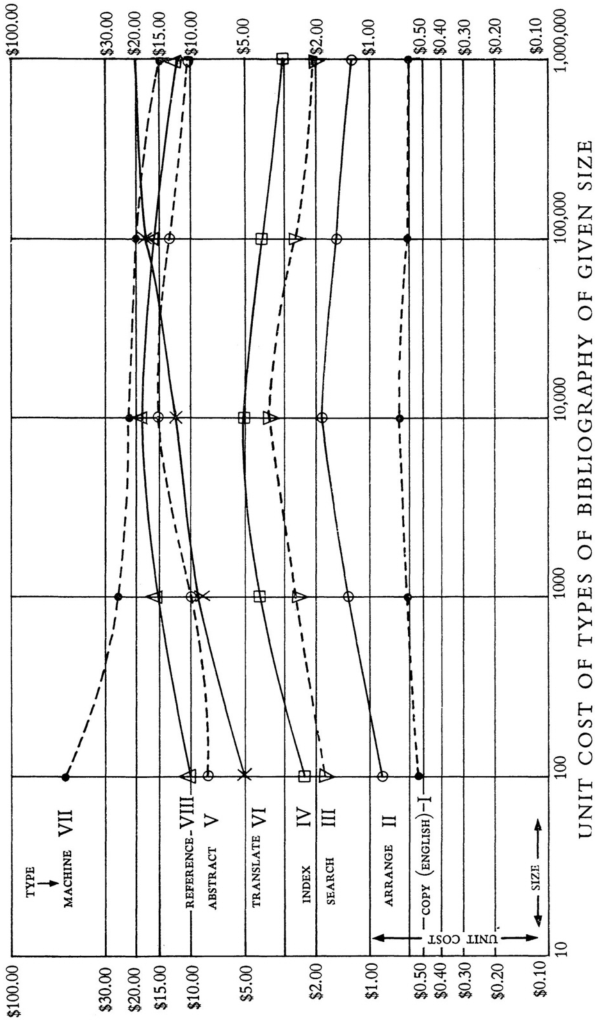


FIGURE 1. Nomogram for bibliographic cost prediction.

which the author has direct knowledge, or for projects with which he is acquainted. Supporting evidence for these figures is contained in the appended "exhibits," which will be available to anyone desiring to verify the approximate figures or the quality of the work done for a given price.

As this is a "first approximation," a great deal of smoothing was necessary. Further study and refinement will reveal interesting deviations from these "order-of-magnitude" curves. Variations of 100% or even 200% due to local factors should be considered normal. Eventually enough data may be available, if uniform analyses are made for dozens of bibliographies of each type, to establish the "Standard Deviation" of some of the values, and certainly some of the curves will need to be relocated. Anomalous increases will occur where increasing complexity due to increasing volume necessitates introduction of new systems of control and development of new tools, codes, class systems, mechanization, searching and procurement methods, and methods of reproduction.

TABLE 2 Total cost of various types of bibliographies of designated size

Type	*Characteristic*	*10^2*	*10^3*	*10^4*	*10^5*	*10^6* [a]
I	Copy	$55.00	$600.00	$7,000	$60,000	Nonsense
II	Arrange	85.00	1,300	18,000	150,000	Nonsense
III	Search	180.00	2,500	36,000	260,000	$2,000,000
IV	Index	230.00	4,000	50,000	400,000	$3,000,000
V	Abstract	800.00	10,000	150,000	1,300,000	$10,000,000
VI	Translate	500.00	9,000	120,000	1,800,000	$20,000,000
VII	Mechanize	5000.00	25,000	220,000	2,000,000	$15,000,000
VIII	Reference	1000.00	15,000	180,000	1,500,000	$12,000,000

[a] These figures for a Sputnik type of over-all abstracting service should probably be doubled to allow for inflation while the project is being discussed, explored, coordinated, approved, financed, activated, and mechanized or put into operation (about 3 to 6 years).

The *eight basic types* represented by the eight curves on the cost prediction nomogram (Fig. 1) are as follows:

I. Straight *copying* from already prepared lists or cards.
II. Copying onto cards, *arranging* and recopying (no searching).
III. *Searching* material from various sources and arranging.
IV. Searching, arranging, and simple *indexing.*
V. Searching, arranging, indexing, and *abstracting.*
VI. Selecting, *translating* foreign abstracts, and indexing.
VII. Searching, arranging, abstracting, and indexing by machine.
VIII. Searching, arranging, abstracting, etc., plus *reference* and bibliographic service on demand from profession.

Subtypes, as shown in the cost analysis form (Appendix 1), take into consideration first, the three types of *language* complexity: (1) straight *English,* (2) *mixed languages* (e.g., English, French, German, Italian, Spanish), and (3) *difficult languages* such as Russian, Hungarian, Finnish, and Japanese, which require employment of specialized and expensive personnel.

Secondly, the degree of *comprehensiveness* or exhaustiveness was considered. It is obvious that it would cost many times more to prepare a very exhaustive bibliography on any given subject than merely to gather together the easily available material.

Thirdly, the variable labor costs and other factors (such as skilled personnel, well-organized analytical catalogs in libraries of countries like England, France, Germany, Holland, Belgium, and Japan, compared with those of the United States, where such analytical catalogs are rare), cheaper reproduction costs, etc., make it about 50% cheaper to do the work abroad if proper arrangements can be made. However, there is some disadvantage in operating by "remote control" and there is also the matter of indicating to the user the availability of the material in local libraries where it may be easily obtained.

SUPPORTING EVIDENCE

The cost prediction nomogram (Fig. 1) and Table 1 and the computed total cost data (Table 2) are based on seven controlled projects (Exhibits 1 through 7) performed under this NSF Grant, and analysis of more or less complete information obtained from *Meteorological Abstracts* projects and from other services doing similar bibliographic work. Copies of the seven controlled bibliographies will be available at the Conference for the benefit of anyone wishing to check the type, quality, or cost of each or all. Detailed breakdowns of the cost figures appear in this paper following Appendix 1.

Exhibit 1 consists of a simple listing of the authors, titles, and serial citations for 100 selected references to articles on Antarctic Meteorology, all in the English language, with no attempt to check, arrange, index, or edit the material copied (Type I, Fig. 1). Duplication (100 copies) was by mimeograph, with a simple title page. Total cost was $55.05 or $0.55 per item. In some cases, and in some countries, the total cost might be half; in other cases 50% higher.

Exhibit 2 is similar to Exhibit 1 except that foreign language material on Antarctic meteorology was included, and the references were copied onto cards, arranged, numbered, and mimeographed (100 copies). No cataloging, annotating, indexing, or editing was involved. Costs amounted to $81 for 100 or $0.81 per item. This is considered as a variation of Type II and is included on the cost prediction nomogram.

Exhibit 3 (an example of Type III) involved detailed searching for references to articles on noctilucent (stratospheric) clouds from 1885 to date, and included many German, Scandinavian, and Russian articles. Most items were checked against the original source, and in no case was there an extensive list of references or card file available for copying Work was done in Washington, D.C. Total cost was $172.50 for 102 items or $1.70 each.

Exhibit 4 is exactly the same type (III) and subject as Exhibit 3, except that the work was done in Brussels, Belgium, by Dr. Vandenplas, an experienced bibliographer who had available in his institution fairly complete and detailed analytical files of references to specialized subjects in meteorology or astronomy. He compiled 195 references for $105 or a total unit cost, including reproduction of 100 copies (in USA), of $0.54 per item. The material was arranged chronologically, but the references were not checked or edited to any extent.

Exhibit 5 is like Exhibits 4 and 3 (Type III), but the work was done in Japan by Dr. H.Arakawa, also an experienced meteorologist and bibliographer. He turned up 156 items in a fairly exhaustive search. Total cost was $105.30 or $0.67 per item. In this case as well as in the one done in Belgium, costs would have been somewhat less if the reproduction had been done in the country of the compiler, for mimeographing is about half as expensive per page in other countries as in the USA. However, this was not possible, since it was desirable to have uniform editing and reproduction.

Exhibit 6 on noctilucent clouds also is of Type III. The work was done in England, searching was fairly exhaustive and included a number of Russian articles and much German, Scandinavian, and other language material. Editing was not as carefully done, but entries were fairly complete. Cost for 224 items was $124.00, or $0.51 per item. Source material in the British Meteorological Office is fairly well organized by subject, because of 50 years of careful analytical work by one man, Dr. C.E.P.Brooks, who was the world's leading documentation expert in meteorology for a quarter of a century. However, some additional material was ferreted out, to make the list exhaustive.

Exhibit 7 (Type IV) is an exhaustive, briefly annotated bibliography of 193 items on works of Dr. C.E.P.Brooks. Most of these works are in English. Cost was about $450 or $2.33 per item, but duplication has not yet been completed. An index was provided.

CONCLUSIONS

The results of this study of unit costs of several different types of bibliographies prepared, under controlled conditions, in the United States and abroad and of a number of more comprehensive bibliographies and abstracting or indexing services reveal the fact that costs can vary by two orders of magnitude (50c to $50 per item) in extreme cases and by one order of magnitude ($1 to $15. per item) in the normal run of indexing or abstracting service.

The variations in cost are due to the factors that are listed and to other unpredictable factors which may be revealed by further study:

1. Amount of bibliographic work done previously by someone else, or some library or bibliographic service.
2. Exhaustiveness of coverage.
3. Amount and type of annotation or quality of abstracting.
4. Amount of foreign language material, especially if in difficult languages.
5. Amount of arrangement or rearrangement necessary.
6. Thoroughness of classification and indexing.
7. Country where work is done (availability and cost of labor).
8. Quality and accuracy of bibliographic entries or citations.
9. Type, quality, and volume of reproduction.
10. Whether or not abstracts are furnished free or by author.

Another important element is the amount of service rendered in addition to the visible product: Development of new methods, tools, reference service, special bibliographies, interagency or international cooperation, printed cards or microcard service to libraries, etc.

The cost prediction nomogram (Fig. 1) and Tables 1 and 2 show the relative order of magnitude of the unit cost, or the total cost of any given type or size of project. These costs will rise with passage of time because of increased labor and printing costs.

APPENDIX 1 AMERICAN METEOROLOGICAL SOCIETY, INTERNATIONAL CONFERENCE ON SCIENTIFIC INFORMATION, BIBLIOGRAPHIC COST PROJECT

Place: ______________________________

Institution: ______________________________

Compiler: ______________ Editor: ______________

Title of Bibliography or Bibliographic Service: ______________________________

____

Type of Bibliography: ______________________________

Sub-types: ______________________________

Size or number of items included: ______________________________

I Total ______________ II. Per year ______________

Cost:

I Total cost ____________

II Unit cost. A. Per 100 items ____________ B. Per item ____________

Cost analysis:

I Organization ____________
II Searching or locating ____________
III Copying or cataloging ____________
IV Arranging ____________
V Classifying ____________
VI Indexing ____________
VII Abstracting or translating ____________
VIII Editing ____________
IX Typing and proofreading ____________
X Duplicating or printing ____________
XI Overhead (materials, administration, etc.) ____________

Types of bibliographies:

I Copying (from lists or cards)
II Arranging and copying
III Searching and arranging
IV Searching, arranging, indexing (simple, subject, systematic, cumulative)
V Searching, arranging, indexing, abstracting (A. Indicative, B. Informative, C. Critical)
VI Selecting, translating and indexing foreign abstracts.
VII Searching, arranging, abstracting, indexing by machine.
VIII Searching, arranging with reference and bibliographic service.

Sub-types:

A. Language factor
   1. English
   2. Mixed languages
   3. Difficult languages

B. Comprehensiveness
   1. Selective
   2. Representative
   3. Exhaustive

C. Country
   1. United States
   2. Abroad. (Specify) ____________
      1. England
      2. Belgium
      3. Japan
      4. Other

D. Type of reproduction
   1. Typed
   2. Mimeographed
   3. Photo-reproduction
   4. Book press

Exhibit 1 *Simple copying of English language material: A selective bibliography of English language publications on Antarctic meteorology*

This bibliography is one of the simplest types of bibliographies, involving the locating of a published bibliography on the subject wanted, selecting 100 English language items, copying these items, drafting a suitable title page and duplicating it by mimeographing process. Little editing was done. No knowledge of foreign languages, cataloging, abstracting, or indexing was involved.

Costs involved were:

I	Organization	$ 5.00
II	Locating and copying	21.00
III	Editing	3.75
IV	Typing and proofreading	9.50
V	Duplicating	10.80
VI	Overhead	5.00
	Total cost	$55.05

Exhibit 2 *Copying English and foreign language references: A selective, multilingual bibliography on Antarctic meteorology*

This bibliography, also one of the simpler types, involved the location of a bibliography on Antarctic meteorology containing foreign as well as English language references, the selection of the 100 items desired, the copying of these items on cards, numbering and typing the bibliography, and mimeographing. Some knowledge of foreign languages was needed, since French, Spanish, German, and Russian entries were included. No cataloging, abstracting, or indexing and little editing were done.

Costs involved were:

I	Organization	$ 5.00
II	Locating and copying	51.00
III	Typing and proofreading	8.50
IV	Editing	2.50
V	Duplicating	9.00
VI	Overhead	7.50
	Total cost	$83.50

Exhibit 3 *Searching, copying, and arranging multilingual references from various sources; work done in the USA: A bibliography on noctilucent and nacreous clouds*

(by Marian K.Rigby)

This bibliography was more complex, though unannotated or indexed. The material had to be searched or located, in most cases cataloged and cards made. Languages involved, besides English, were Italian, German, Russian, French, Norwegian, and

Hungarian. The 100 odd cards were alphabetically arranged, then typed, and finally mimeographed. A great deal of time was required in searching and cataloging, since few of the items had been previously included in any catalog. This also was quite exhaustive.

Costs involved were:

I	Organization	$12.50
II	Searching or locating	40.00
III	Copying or cataloging	64.50
IV	Arranging	4.50
V	Editing	12.00
VI	Duplicating	30.00
VII	Overhead	9.00
	Total cost	$172.50

Exhibit 4 *Searching, copying, and arranging multilingual references from various sources; work done in Belgium: Noctilucent and nacreous clouds*

(by DR. A.VANDENPLAS)

This bibliography is similar to the previous one, but it was compiled and typed in Belgium. It is arranged chronologically, but not indexed, classified, or annotated. It was compiled from previously cataloged cards and contains 195 items.

Costs involved were:

I	Organization	$10.00
II	Searching, locating, copying, and arranging	32.40
III	Editing	3.00
IV	Duplicating	53.70
V	Overhead	6.00
	Total cost	$105.10

Exhibit 5 *Searching, copying, and arranging multilingual references from various sources; work done in Japan: Noctilucent and nacreous clouds*

(by DR. H.ARAKAWA)

Compiled in Japan, this bibliography of 156 items also resembles the one done in the United States. It is arranged chronologically and was fairly exhaustive, but it is unannotated, classified, or indexed, and required little editing. Mimeographing was done in the United States.

Costs involved were:

I	Organization	$10.00
II	Searching, locating, copying, and arranging	38.30
III	Editing	3.00
IV	Duplicating	48.00
V	Overhead	6.00
	Total cost	$105.30

Exhibit 6 *Exhaustive searching, copying, and arranging multilingual references from various sources; work done in England: A bibliography on noctilucent and nacreous clouds* (by R.BAKER)

This bibliography, of 224 items, on noctilucent and nacreous clouds, was compiled in England and is the most exhaustive of the four. It is arranged chronologically, is unannotated, unclassified, and is not indexed. It required a good deal of editing.

Costs involved were:

I	Organization	$10.00
II	Searching, locating, copying, and arranging	(50.00) (est.)
III	Editing	9.00
IV	Duplicating	49.00
V	Overhead	6.00
	Total cost	$124.00

Exhibit 7 *Exhaustive searching, annotating, arranging, and indexing mostly English language material: A bibliography of the works of Charles Ernest Pelham Brooks*

This is an exhaustive bibliography, is completely annotated, arranged chronologically, indexed by subject and geographically. Before his death, Dr. Brooks edited it himself, and in many cases provided the abstract and classification.

Costs involved were approximately $450.00.

The Efficiency of Metallurgical Abstracts

NERIO GAUDENZI

To discuss the efficiency of metallurgical abstracts, after what was said by Frank T.Sisco at the 1957 Annual Meeting of the National Academy of Sciences, may appear as an absurdity, or at least as presumptuous. In accordance with the statements of the Director of the Engineering Foundation of New York, it probably would be advisable to speak about a failure of metallurgical documentation, and of documentation in general in the abstracting field, if, as Mr. Sisco said, "If a research job in the USA costs less than $100,000, it is cheaper to do it than to find out if it has been done before and is reported in the literature." (5).

But while we wait for results, which may be available in 1960, of five-year studies, promoted by the ASM in 1955, in the pilot plant at the School of Library Science at Western Reserve University, to determine if machine methods of searching and finding coded metallurgical literature are practical, it will be not useless to point out the actual situation.

It may be that in America automation, when introduced in the metallurgical literature searching field, will produce revolutionary results, or at least results more satisfactory than those achieved up to now with usual methods. In Europe, however, the conditions are quite different, and probably it still will be prudent for many years to try to improve the defects of existing documentation systems.

Our statements, which are the fruit of direct researches published previously, (4), are the result of work carried out by one who is both a producer and a user of abstracts in the field of metallurgy. They confirm or support the candid but substantially exact evaluation of Mr. Sisco.

We must recognize first that technical documentation has had in these last years a considerable development and that, therefore, there exists at the present time abundant material for studies which, unfortunately, being spread in many streamlets, are still very little known and are not adequately used. Standards and prescriptions resulting from careful studies in the field of bibliographical

NERIO GAUDENZI Editor-in-Chief of the journal *Alluminio,* ISML, Milan, Italy.

abstracting are very often ignored by those who should apply them. Recommendations formulated during international meetings and by qualified associations, and prescriptions elaborated by standards institutions often remain a dead letter, and old lines already outmoded by more recent requirements continue to be followed. This imbalance is due also to the lack of a bridge between the researcher and the documentalist. The bridge theoretically is built by the technical documentalist, who in Italy usually is a graduate in engineering, in chemistry, or in the physical or mathematical sciences and who, owing to casual circumstances, is obliged to deal with documentation problems; or he may be a graduate of the faculty of arts who, always for professional reasons, has to deal with technical documentation, which is looked upon as a part of general documentation. The first, coming from the technical-scientific faculty, is well acquainted with disciplines he is graduated in, but is not sufficiently a master of the documentation techniques, since the knowledge acquired in the university in this field is absolutely inadequate. The second, coming from faculties of arts (philosophy and languages), might well have the mental structure and fundamental characteristics of a bibliographer but, having no specific knowledge of the subject, cannot have adequate competence to resolve the concrete problems of documentation as applied to physics, chemistry, engineering, etc.

Let us consider now the documentation and especially abstracting in the field of metallurgy. This consideration could serve as an example of combining the requirements of the documentalist with those of the researcher in the field of metallurgy.

The problem of the efficiency of bibliographical abstracts in metallurgy, as well as in any other branch of science, can be examined from a double point of view, i.e., quantitatively and qualitatively.

QUANTITATIVE EFFICIENCY

Apparently quantitative efficiency is not difficult to evaluate. Theorically, it is sufficient to know the number of journals of metallurgy published in a year and the number of articles contained in them. If examination of bibliographical abstract journals shows that *all* articles published have been abstracted, we can say that the network of abstracts is perfect, since not a single article has escaped abstracting. The more the system deviates from this theoretical and ideal condition the lower is the efficiency of the service.

Let us see what the situation is in practice. First of all, it is necessary to decide what we mean by a journal of metallurgy. We can use the definition adopted in our previous study (4, p. 15), which naturally is subject to criticism, on the basis of which metallurgical journals were considered to be “those journals whose titles contain a clear reference to mining and metallurgy treated as a science and including everything from ore to finished products."

On the basis of long years of study on this subject, we believe that it is possible to calculate that there are, in the entire world, approximately 2000 metallurgical journals actually "alive." This number can be taken as a starting point, but it is not complete; many articles of a metallurgical character are being published in journals of chemistry, physics, engineering, etc. For example, results from our earlier study (4, p. 32) show that in the field of light metals, in 1954, of the articles abstracted in various abstracting journals, 34% were reported from nonmetallurgical journals. This scattering of publications on a certain subject (in our case metallurgy) throughout an enormous number of journals ought to be reduced by authors to a necessary minimum. Naturally, in the case of metallurgical publications it is impossible to avoid the dispersion as, for example, an article on the use of aluminum in the textile industry belongs in a textile journal rather than in a metallurgical magazine, but articles on the science of metals or the technology of metallurgy should be published in journals of metallurgy. Also prompt notice of original articles published could be given by authors or by editors of journals to a central bibliographical association, charged with the collection of such data.

In any case, based on the actual situation, if it is assumed that for various abstracting services all 2000 metallurgical journals will be examined, about a third of the articles published would escape and the efficiency would be only 70%.

If we next consider the number of bibliographical abstract journals concerned with metallurgy, we can distinguish two types of publications: *(a)* those dealing *exclusively* with metallurgy, e.g., *Metallurgical Abstracts* of the Institute of Metals, or *Extraits* of the *Revue de Métallurgie* and *(b)* those having a general technical character and containing sections devoted to metallurgy, e.g., *Chemical Abstracts, Chemisches Zentralblatt, Bulletin Signalétique du Centre National de la Recherche Scientifique.*

The first group of publications number about 70, if we judge from our first study (4, Appendix 1), based on the examination of publications existing on this subject and, above all, on the direct control of journals. In the second group, in which we naturally include only abstracting journals of wide coverage and completeness, can be mentioned the following: (1) *British Abstracts. BI Chemical Engineering and Industrial Inorganic;* (2) *Chemical Abstracts;* (3) *Engineering Index* (in English); (4) *Chemisches Zentralblatt* (in German); (5) *Bulletin Signalétique du Centre National de la Recherche Scientifique* (in French); (6) *Boletin (del) Centro de Documentacion Cientifica y Tecnica, Mexico* (in Spanish); (7) *Referativnyi Zhurnal Metallurgia* (in Russian). The total number of abstracting journals dealing with metallurgy can be taken as about 80.

If we now wish to establish a relation between original metallurgical journals and those publishing bibliographical abstracts of metallurgy, we find that this is about 2000:80, i.e., one abstract journal to every 25 original journals. This ratio seems normal because, according to available data contained in the literature, there should exist actually about 50,000 original technical and scientific[1] and about 1400 abstracting journals (3), with a ratio of 35.

An outsider might think that such a high number of abstracting journals in metallurgy might mean a higher efficiency since there would be a better guarantee that all articles published would be covered. Anyone who has had experience in this field, however, knows that the efficiency is quite limited because every journal operates independently; consequently, many articles are abstracted several times, while others are missed altogether. For example, in the field of light metals we learned in our previous study (4, p. 30) that some journals (*Modern Metals,* Chicago; *Light Metals,* London; *Aluminium,* Düsseldorf; *Light Metal Age,* Chicago; *Alluminio,* Milan) are covered by 10 to 12 abstracting services; the total almost certainly is at least 20, since our previous investigation was limited to specialized abstracting journals and the 5 primary publications mentioned above are regularly abstracted also by abstracting journals having a general character.

In our opinion a satisfactory solution of the quantitative problem could be put on the following basis. Limit the number of abstracting journals to principal fields offering a specific interest for certain groups of industry or institutions, as it appears for example, from the following: (1) one of general character (science of metals), including also the field of mining and metallurgy; (2) one dedicated to iron metallurgy; (3) one to nonferrous metals; (4) one to light metals; (5) one to foundry technology; (6) one to welding; (7) one to corrosion and surface treatments; and (8) one to economics and statistics of metallurgy. Naturally, there should be as close coordination as possible to minimize duplication. Actually, the picture of metallurgical abstracting journals is as follows (4, p. 27):

1. Science of metals	17
2. Iron metallurgy	7
3. Nonferrous metals	9
4. Light metals	14
5. Foundry	3
6. Welding	6
7. Corrosion and surface treatments	8
Total	64

[1] The Academy of Sciences of the USSR (6) examines about 12,000 journals, of which about 9700 are printed in Latin alphabet types, about 1900 in languages having Cyrillic alphabets, and the rest (about 400) in languages with other types. Among the group with Latin types, only about 300 are in metallurgy.

It is obvious that a considerable reduction in the number of these publications should be possible.

Every journal should carry abstracts in 5 principal languages: English, French, German, Russian, and Spanish. Naturally, where there is a qualified international institution for intercommunication, a unique journal could contain abstracts in all 5 languages or in some of them (as is done by the International Welding Institute or the *Bibliographic Bulletin for Welding and Allied Processes* which is published in French and English).

In other cases, a unique journal in a determinate language could deal very well with all eight subjects, as is done today in Russia where a centralized organization permits such an enterprise. The *Referativnyi Zhurnal,* which standardizes the abstracting service for all sciences and is divided into separate sections for each of them, has achieved a considerable degree of completeness. Its metallurgical section published 14,957 abstracts in 1956, whereas the most advanced service in the free world *(ASM Review of Metallurgical Abstracts)* issued only 8000. A most laudable enterprise has been announced this year by *Acta Metallurgica*—the publication in English of abstracts appearing in the part *of Referativnyi Zhurnal, Metallurgy* which covers original journals that are very difficult to find in western countries.

The number of bibliographical journals of metallurgy could be reduced to 5, if based on a linguistic principle, and to 8 if based on the principle of contents, to reach the maximum of 40. At least 40 of the existing journals could then be eliminated, and the service, with minor dissipation of forces, could be improved.

To improve the efficiency of metallurgical abstracts it will be necessary, from our point of view, to centralize the service under a single institution in order to avoid a duplication of work and to assure coverage of areas now neglected. In this connection, attention could be called to an interesting idea actually under study by CIDA (International Centre for the Development of Aluminum) with headquarters in Paris, of which the following countries are members: Austria, France, Germany, Great Britain, Italy, Sweden, and Switzerland. This organization has in mind to centralize the abstracting service in the light metals field, and we hope this will be realized soon and concretely.

International cooperation in the field of bibliographical abstracting is indispensable, particularly in Europe, where, since funds for this kind of work are scarce, they must not be wasted. On the other hand, the less well-off countries must have access to information on the research carried out in richer countries, so that for the former documentation assumes a greater importance than for the others.

QUALITATIVE EFFICIENCY

Let us now examine the qualitative efficiency of bibliographical abstracts. As is known, there are today two types of bibliographical abstracts—indicative and informative.

According to the definition of Unesco (2), "an *indicative abstract* is a short abstract written with the intention to enable the reader to decide whether he should read the original publication or article"; that is, this kind of abstract indicates only what an article is about without attempting to be a substitute for it. The searcher, if the subject is of interest to him, must himself procure the original publication.

The *informative abstract* "summarizes the principal arguments and gives the principal data contained in the original publication." In this case the abstract may substitute for the original article and the searching of the latter can be avoided.

This distinction, theorically quite precise, actually is a complicated one in practice. For one thing, the compilers of the abstracts do not always have these definitions clearly in mind. Also, compilation of informative abstracts requires a special competence in the material and a knowledge of preceding literature, which only a specialized person can have. For example, in effect, only one who deals with and lives in the foundry atmosphere is able to compile really informative abstracts on this part of metallurgy.

This is what happens to the compiler of abstracts, while the users and documentalists, who are not always specialists, will find it difficult to judge at once whether an abstract is an indicative or an informative one, except in limited cases. No documentalist, for example, will be in doubt about the type of abstracts (informative) published in the *Extraits* of the *Revue de Métallurgie,* which each take one or more pages of the journal, or about the indicative ones of the *Metals Review* of the ASM, which consist of only a few lines. In this case, the criterion of method of compiling abstracts may be linked with length: an informative abstract is normally extensive and an indicative abstract is short. Thus in practice it happens too often that length is used as the basis of discrimination between indicative and informative abstracts, this being a principle that is much easier and more comfortable to apply than one based on the method of compilation. We have made these points to prove that it is not easy to establish which type of abstracts is practically preferred in the field of metallurgy.

From the direct examination of 56 abstracting journals (4, Appendix 1), using evaluations devised by us which do not always coincide with those of other bibliographical researchers, we concluded that 39 of the journals have adopted indicative abstracting, 20 journals informative, and 31 journals semi-informative.[1] This last group includes abstracts of medium size, compiled according to a criterion between the informative and the indicative, constituting a neutral zone where assignment as to the one or the other principal type of abstracts would be highly objective. The statistical principle does not furnish sufficient information, and this lack is not eliminated through taking the language as a basis.

Of French journals, 19 carry indicative, 7 semi-informative, and 6 informative abstracts. In journals printed in the English language the principle of semi-informative abstracts prevails in 14, indicative in 12, and informative in 10. In the German language 8 journals are semi-informative, 4 are indicative, and 3 are informative. In journals in Italian and other languages all three types of abstracts have the same number—one.

If, however, we confine our interest to abstracting journals of more common use and of greater diffusion, known by a great many researchers and documentalists, we will perhaps find more reliable information.

In France, for example, we can clearly recognize one tendency, as the *Revue de Métallurgie,* official publication of the Société Française de Métallurgie, the *Circulaires d'Information Techniques* of the Centre de Documentation Sidérurgique, *Fonderie,* and the *Documentation Métallurgique,* are issuing quite long abstracts which are decidedly informative in character. The use of indicative abstracts is limited to the *Bulletin Signalétique* of the Centre National de la Recherche Scientifique.

In the United States, on the other hand, the most important metallurgical abstracting service, that of the American Society for Metals, which publishes its abstracts in the monthly *Metals Review* and then republishes them in annual volumes of *ASM Review of Metal Literature,* has adopted the indicative form of abstracting. Furthermore, if we were able to find 10 informative abstracting journals, they would not be too extensive and would be closer to the semi-informative than to the strictly informative type.

The abstracts of NACE which, although they are published both in the monthly journal *Corrosion* and in the *Annual Bibliographic Survey of Corrosion,* also tend to the indicative form of abstracting.

If we further consider the *Metallurgical Abstracts* of the Institute of Metals, London (also published monthly in the *Journal of the Institute of Metals*) and the *Abstracts of Current Literature* (published in the *Journal of the Iron and Steel Institute,* London) we find that both of these have a semi-informative character, tending, as to length, more toward the indicative type of abstracting than to the

[1] The sum of 39+21+31 is superior to the number of journals examined (56) for many journals have adopted more than one type of abstract.

informative. Thus, it is obvious that in English-speaking countries (United States and Great Britain) preference is being given to an indicative or, at least, to a semi-informative type of abstracting in the metals field.

In Germany, a country in which bibliographical services are particularly well developed and appreciated, there is no bibliographical journal of metallurgy of general character, and one is obliged to consult the classical *Chemisches Zentralblatt,* corresponding to *Chemical Abstracts* in the United States.

In the field of iron metallurgy there is the *Zentralblatt der Hütten-und Walzwerke,* included in the journal *Stahl und Eisen,* which is decidedly indicative in character; also numerous other journals devoted to particular fields of metallurgy have, in general, a semi-informative character, tending, however, more to the indicative than to the informative form.

If we depart from the difference between indicative and informative abstracts, it is evident that in English-speaking countries there exists a better and more complete bibliographical organization as compared with that in France or in German-speaking countries.

We now shall pass on to the examination of other elements which could be advantageous for establishing the value of abstracting services and of the abstracts themselves.

In order not to digress too much from the main subject, we shall consider only the principal aspects that have general bibliographical value and interest, including particularly abbreviations of the titles of the abstracted journals, the coding systems adopted by various services, and the manner of presentation of abstracts (pages of journals or cards).

ABBREVIATION OF THE TITLE OF THE JOURNAL EXAMINED

Where, because of space limitations, it is not convenient, in the abstracts, to reproduce the full title of the original publication containing the article abstracted, it will be advisable to use the following:

1. DIN 1502, Germany, 1955 (1st edition 1940).
2. Recommendation ISO R 4 of international character, May 1954 (inspired by the ISA Bulletin 23 of 1940).
3. Code for Abbreviations of Titles of Periodicals, of 1953, Great Britain.
4. AFNOR 2–44–002 of 1944, in France.

In Belgium, India, Holland, Denmark, Spain, and Switzerland there are similar standards.

In the field of metallurgy, judged by the results of an investigation carried out by us, *no one* of the services has shown any interest in following these standards, and abbreviations are still made in a completely arbitrary way. For example, speaking only of the more important services,[2] we find that one and the same title is abbreviated in many different ways, as:

1. ASM Review of Metal Literature: *Metal Treatment and Drop Forging*.
2. Journal of the Iron and Steel Institute: *Met. Treatment*.
3. Light Metals Bulletin: *Metal Treatment & Drop Forging*.
4. Metallurgical Abstracts: *Metal Treatment*.
5. Stahl und Eisen: *Metal Treatm. Drop Forg*.

If, instead, we consider abstracting journals of general character, having sections on metallurgy, we find for the same title the following abbreviations:

6. Chemisches Zentralblatt: *Metal Treatment Drop Forging*.
7. Chemical Abstracts: *Metal Treatment and Drop Forging*.
8. Bulletin Signalétique du Centre National de la Recherche Scientifique: *Metal Treatment*.
9. Referativnyi Zhurnal: *Metal Treatm. and Drop Forging*.

It is difficult to determine why there has been no standardization of abbreviations in the field of metallurgy. It may be that the compilers of various services (especially the smaller and more specialized ones) simply ignore the matter or consider it not worth the trouble. It may also be true in the case of more important services, which began their publications before there was any standardization, that for practical reasons they do not want to change the system already adopted by them; also it may be that the standards themselves are not as easy to apply here as with ISO R 4. This document prescribes (paragraph 4), “The normal method of abbreviation consists in suppressing the last letters of the word, at least two, substituting them with a full stop.” As for example the word “Journal” can be abbreviated as Journ., Jour., Jou., Jo., J.; the possibility of five different solutions in one standard seems to permit excessive freedom of interpretation, contrary to the very spirit itself of standardization.

CODING SYSTEM

The coding problem is highly difficult and complex in the field of bibliography and a system which can satisfy everybody probably can never be found. In this review, we would like to consider only what principles have been adopted in practice for bibliographical abstracts in the field of metallurgy. Above all, it is necessary to distinguish between the real classification of general validity and

[2] We have adopted the principle of Varossieau (1) who indicates as important those bibliographical services which publish at least 4000 abstracts in a year.

a generic classification. To the first group belong, for example, the decimal system and the ASM-SLA. In this case every article receives a fairly long symbol, formed by letters or by numbers, which identifies individual articles and affords the searcher the possibility of forming a classification which represents a permanent bibliographical consultation source, in which every subject will be kept constantly up to date. Under these circumstances, no importance is to be attributed to the fact that the classification has a universal value (at least in theory), i.e., whether it can be used for all sciences and by all systems, or whether, on the contrary, its value is limited to a specific field like the ASM-SLA system for metallurgy. The main thing is that it will be possible with the classification to maintain constantly up-to-date files in which every single article will be easily traceable.

There are other less subtle coding systems which we can call generic. These are limited to the constitution of comparatively wide subdivisions, in which all articles on related subjects find their places, as in our field, welding, corrosion, and finishing. This principle, as we shall see, is generally applied. The *Metallurgical Abstracts* of the Institute of Metals, to indicate one of the most important and authoritative abstracting services, follows this system. Articles are subdivided as follows: (1) properties of metals; (2) properties of alloys; (3) structure; (4) dental metallurgy; (5) powder metallurgy; (6) corrosion and related phenomena; (7) protection; (8) electrodeposition; (9) electrometallurgy and electrochemistry; (10) refining; (11) analysis; (12) laboratory apparatus, instruments, etc.; (13) physical and mechanical testing, inspection, and radiology; (14) temperature measurement and control; (15) foundry practice and appliances; (16) secondary metals, scrap, residues; (17) furnaces, fuels, and refractories; (18) heat treatment; (19) working; (20) cleaning and finishing; (21) joining; (22) industrial uses and applications; (23) history of metals; (24) metal economics; (25) health hazards; (26) miscellaneous; (27) bibliography; (28) book reviews.

With this system, retroactive searching of abstracts naturally becomes increasingly laborious as the years pass. With this type of classification, the bibliographic research of a specific subject becomes very long, because it is necessary to examine all articles on related subjects (for example, for a journal of forging one would have to consult all items on working, year by year).

Beyond the above, there are other classification systems—if they can be called such —which give only a numeration, generally progressive per annum, of articles abstracted: this has, above all, a practical value, as it can be used for a quick marking of abstracts by users of abstracts who are looking for an original article.

Based on these principles, and with data previously collected by us (4, Appendix I), we observe that among 53 metallurgical abstracting services:

33 have adopted a generic classification system
7 have adopted the decimal classification system
6 have no classification system
5 have their own classification system
2 have adopted the ASM-SLA classification system[3]

If the examination is limited to important services, we find that only one *(ASM Review of Metal Literature)* has adopted the ASM-SLA classification system, while all others content themselves with classifications of generic character.

These data show that in the field of metallurgy, no one of the important services actually is using the decimal classification system. Among the less important services, 7 have adopted the decimal system. They are:

1. AGM Leichtbau (light metals), Germany	700 abst./year
2. Literaturbericht (light metals), Switzerland	500 abst./year
3. Archiv für Metallfinishing (surface treatments), Germany	300 abst./year
4. Hutnik (general metallurgy), Poland	100 abst./year
5. IMM Abstracts (mining & metallurgy), England	2800 abst./year
6. SMRE Abstracts of Current Publications (safety of mines), England	1000 abst./year
7. Zeitschrift für Metallkunde, Germany	—

Of these, some are not real bibliographical journals (1, 3, 4, and 7); No. 2 is more like a restricted bulletin, and only the 5th and 6th can be said to be important from the documentation point of view in the field of mining. Therefore, it can be stated that in effect the field of metallurgy (excluding mining) has practically completely abandoned the decimal classification system.

Instead of the decimal system, the ASM-SLA is being used by ASM, which is its promoter, and by AIM (Italian Society of Metallurgy) for its *Atti e Notizie AIM*. Other special classification systems are being adopted, like that of the NACE in *Corrosion,* of the Istituto Internazionale della Saldatura in the Bibliographical Bulletin for Welding, that of the Centre de Documentation Sidérurgique in their *Circulaires d'informations Techniques* and *Bulletin Analytique,* and those of the Documentation Métallurgique and of the Svetsliteratur.

[3] According to information we received from the AIM, French and German iron metallurgy associations also have decided to adopt the ASM-SLA classification system.

The majority of the services use instead more or less similar classification systems.

PRESENTATION OF ABSTRACTS

The presentation problem of abstracts, i.e., whether to publish abstracts on cards or on normal pages has considerable importance from the documentation point of view. The first type presumably should be preferred by the searcher of abstracts, as it permits maintenance of both author and subject files, which can be kept continuously up to date and consulted rapidly and easily. For this the user naturally must have a documentation office with specialized personnel —something not always possible for small societies. On the other hand, the card system is more difficult and more expensive for the producer of abstracts, who must have an extra service, the cost of which is almost never compensated for by requests of users.

Our earlier study (4, Appendix I) showed that of 63 journals only 3 insert leaflets printed only on one side which can be detached and inserted on white cards to form a file. These are *Aluminium,* Berlin (about 150 abstracts/year), *Documentation Métallurgique,* Paris (500 abstracts/year), and *Ciencia y Tecnica de la Soldadura,* Madrid (500 abstr./year).

Ordinarily, preference is given to presentation on regular pages of journals.

At this point the punch card system must be remembered, which is assuming a considerable development in recent years, and has one of its concrete applications in the documentation of metallurgy. Since 1950, the NACE, Houston, Texas, has offered a regular subscription service of punch cards, containing abstracts of previously published articles; about 2000 cards are issued in a year, at a cost of $100. Also ASM is distributing punch cards that follow its own classification system. These are blank and the user must attach the abstracts.

It must be noted that these two institutions parallel their card services with a continuing conventional service of abstracts presented in the normal form in their periodicals *Corrosion* and *Metals Review,* as well as in their respective annual publications.

As to this use of punch cards, we believe that this is a very important improvement, though at the present time it means a certain expense for the user. The annual cost of NACE punch cards is about $100, while the corresponding volume, *Bibliographic Surveys of Corrosion,* costs $10 (members).

The practical diffusion of this punch card system will tell us whether or not its advantages more than offset its higher cost; our present opinion is that they do.

CONCLUSIONS

From these considerations, it appears clearly that the relatively low efficiency of metallurgical abstracts is due above all to the excessive number of abstracting publications which operate without any connection with each other, so that many original journals are examined by the same services, whereas other original journals escape the network of abstracts entirely. The international bibliographical standards are scarcely observed. With these limitations, it can be observed that English-speaking countries have better services, and that the most complete service belongs to Russia.

The efficiency could be improved appreciably: (1) by centralizing abstracting services, which could be carried out, for the entire field of metallurgy or for single branches (light metals, welding, etc.), by institutions which already have adequate installations, with the cooperation within international orbit (initiative of CIDA), outside of government or bureaucratic interference; (2) by reducing the number of abstracting journals, which actually is excessive; (3) by studying efficient systems able to reduce the dispersion of articles of metallurgical character in journals belonging to other fields; and (4) by distributing abstracts on punch cards.

REFERENCES

1. VAROSSIEAU, W.W. A survey of scientific abstracts and indexing services. *Rev. Document 16,* 25–48 (1949).
2. ——, *International Conference on Science Abstracting. 20–25 June 1949.* Unesco, Paris, 1951.
3. BESTERMAN T. *Index Bibliographicus, Directory of Current Periodical Abstracts and Bibliographies. Vol. I, Science and Technology,* 3rd edition, Unesco, Paris, 1952.
4. GAUDENZI, N. I riassunti bibliografici nel campo metallurgico con speciale riguardo ai metalli leggeri: analisi critica. *Quaderni monografici ISML no. 1.* ISML, Milano, 1956.
5. SISCO, F.T. What's in the literature. *Metal Progress,* 122–124, 1957.
6. Institute of Scientific Information of the USSR Academy of Sciences, *List of Abbreviated and Full Titles of Technical and Scientific Publications.* Academy of Sciences of the USSR, Moscow, 1957.

Subject Slanting in Scientific Abstracting Publications[1]

SAUL HERNER

In recent years, as a result of rising publishing costs and a general increase in scientific publication, the burden placed upon scientific indexing and abstracting publications has mounted significantly. This increased burden on indexing and abstracting publications and their growing inability to cover the scientific literature adequately have given rise to discussions of the feasibility of cooperative abstracting and the common use and exchange of abstracts among the various indexing and abstracting agencies. Various assemblages, including the Royal Society Scientific Information Conference, the Unesco International Conference on Science Abstracting, and, more recently, a meeting of the major American abstracting publications, have given careful and detailed consideration to inter-agency cooperation.

One question which arises in connection with the feasibility of the common use and exchange of abstracts is whether abstracts in the various abstracting publications differ so greatly in structure, content, and emphasis as to necessitate independent preparation, or whether they are similar enough to permit cooperation or exchange. The present paper is a discussion of the results of a detailed analysis and comparison of the abstracting treatment given to representative papers from 51 prominent scientific periodicals by nine major indexing and abstracting publications.

METHOD OF EXECUTION

Since the purpose of the study was to ascertain similarities and differences in the manner in which a given paper is abstracted in the several indexing and abstracting publications, the first step was to obtain a representative sample of papers with a great likelihood of having been abstracted two or more times.

SAUL HERNER Herner and Company, Washington, D.C.

[1] The work discussed in this paper was done under Agreement NSF-G3945, with the National Science Foundation.

This was done by selecting a random sample of 516 papers from the 51 most-cited periodicals in Brown's study of the most frequently cited serials in mathematics, physics, chemistry, geology, physiology, botany, zoology, and entomology. (Cf. C.H.Brown, *Scientific Serials,* pp. 71–142, Association of College and Reference Libraries, Chicago, 1956.) The 51 journals studied are given in Appendix I. By taking a random sample from the most-cited journals on Brown's list, it was possible to ensure that the papers selected for study were representative of their fields, and that they were in journals with a high probability of being abstracted in two or more abstracting publications.

In the selection of papers for study, the January, 1955, issue of each of the sample journals was used. Where publications were issued more frequently than once a month, all issues for the month of January, 1955, were used. Where a given journal was issued less frequently than once a month, and where there was no January, 1955, issue, the first available issue for 1955 was used. The year 1955 was selected as being the latest in which there was a reasonable certainty that the annual indexes to the abstracting publications had been published. The month of January was selected because it permitted the greatest possible amount of time during the year for the abstracting publications to have picked up the papers in the sample. As it developed, it was necessary to turn also to the 1956 issues of the nine abstracting publications in order to obtain a full definition of the abstracting treatment given the 516 sample papers.

The selection and recording of the sample papers was done as follows. A count was made of the total number of articles in the test issues of the 51 specimen journals. This number was divided by 500, the total number of papers required for a satisfactorily small sampling error. The resulting interval between papers was four. The first paper in each issue and every fourth paper thereafter were selected. Descriptive details were then recorded for each paper. These details included the author or authors, title, journal, volume, page, date, language, and the presence or absence of an author abstract or summary.

Each paper was checked against the 1955 and 1956 indexes of the nine indexing and abstracting publications under study, which were as follows: *Applied Mechanics Reviews, Biological Abstracts, British Abstracts of Medical Sciences* (now *International Abstracts of Biological Sciences*), *Chemical Abstracts, Electrical Engineering Abstracts, Geological Abstracts, Mathematical Reviews, Nuclear Science Abstracts,* and *Physics Abstracts*. These abstracting publications were selected as being the foremost in the eight subject fields of the journals whose papers were to be analyzed, and because each was thought to cover publications in two or more of the subjects under study.

The check of the specimen papers against the indexes of the nine abstracting publications produced a list of 207 papers which had been indexed two or more times. These papers are listed in Appendix II. A detailed descriptive record was made of the abstracting of each of the multi-abstracted papers. Included in the information recorded in each case were the following: name of abstract journal, presence or absence of an abstract, whether the abstract was informative or indicative, whether the abstract was a signed or an author abstract, and any distinguishing characteristics which would indicate the degree and character of subject slanting in the abstract. In addition to the descriptive records which were kept for the abstracts in the abstracting publications, analyses were made of all author abstracts or summaries appearing in the original papers. These author abstracts or summaries were given the same descriptive treatment as the abstracts in the abstracting publications. In all, a total of 483 abstracts from the abstracting publications and 152 author abstracts or summaries were examined. This made a grand total of 635 abstracts or author summaries.

CONTENT AND CHARACTERISTICS OF ABSTRACTS

Of the 483 cases in which the specimen papers were picked up by more than one of the nine abstracting publications, 457 gave abstracts and 26 gave only reference citations. The bulk of the 26 items that were given only reference citations were letters to the editor, notes, communications, and other short types of papers.

Of the 457 cases in which papers were actually abstracted, 385, or 84 per cent, received informative abstracts. *Biological Abstracts* and *British Abstracts of Medical Sciences* had the largest proportion and *Mathematical Reviews* had the smallest proportion of informative abstracts. (For the purposes of the present study, informative abstracts were defined as those giving results.) Only three of the 457 abstracts were critical, in the sense that they contained abstractor's comments about the paper being summarized.

Sixty-seven per cent of the abstracts were labeled as original abstracts, and were signed or initialed accordingly. The remaining 33 per cent were labeled as author abstracts. In the case of *Biological Abstracts,* a high proportion of the author abstracts were not taken directly from the papers but were prepared especially by the authors for *Biological Abstracts*.

DISTINGUISHING CHARACTERISTICS OF ABSTRACTS

In order to determine the distinguishing characteristics of the abstracts under study and to detect the presence or absence of subject slanting, all the abstracts of each specimen paper were reproduced and mounted on a single sheet of paper, along with the abstract or summary that appeared in the paper, in the event that there was such an abstract or summary. All sets of abstracts for the specimen papers have been retained for further study, and for perusal by other investigators.[2]

In the course of the preparation of the sets of abstracts, it became strikingly evident that there was a strong resemblance between the abstracts or summaries that appeared in the specimen papers and the abstracts in the abstracting publications. This resemblance prevailed regardless of whether the abstracts in the abstracting publications were marked as author abstracts or whether they were signed by abstractors other than the authors of the papers.

There were, in all, 207 cases in which signed or initialed abstracts were prepared for specimen papers containing abstracts or summaries, and in which the names or initials of the abstractors were not those of the authors of the papers. From a careful comparison of these abstracts with the abstracts or summaries in the papers, it developed that 46 were verbatim copies of the author abstracts or summaries, and 119 were variations of the author abstracts or summaries. There were 42 cases where the signed or initialed abstract bore no clear relationship to the author abstract or summary.

The abstracts below illustrate the comparative abstracting treatment given a specimen paper having an author's summary. The paper, by L.L.Campbell, is entitled, “Purification and Properties of an Alpha-Amylase from Facultative Thermophilic Bacteria.” It was picked up by *Chemical Abstracts, Biological Abstracts,* and *British Abstracts of Medical Sciences*. All three abstracts were signed, and gave the appearance of original abstracts. One was signed by the author, who had prepared a special abstract for the abstracting publication involved. Following is the summary that appeared in the original paper:

Conditions for the production and purification of alpha-amylase from a strain of *Bacillus coagulans* and a strain of *Bacillus stearothermophilus,* following growth at 35 and 55°C. have been determined. The procedure followed for the purification of the enzyme preparations resulted in a 450–600-fold increase in enzyme activity.

A comparison of various properties of the amylase preparations showed that the only significant difference in the preparations produced at 35°C. and those produced at 55°C. was in the thermal stability at 90°C. The 55°C. preparations showed only a 6–10% reduction in activity after 1 hr., whereas the 35°C. preparations showed a 90–92% reduction in 1 hr.

Following is an abstract that appeared in one of the three abstract publications:

Methods for the production and purification of alpha-amylase from a strain of *Bacillus coagulans* and *B. stearothermophilus* following growth at 35 and 55°C. were determined. A comparison of various properties of the amylase preparations shows

[2] These will be available at the Conference for anyone who wishes to see them.

that the only significant difference in the enzyme produced at 35°C. and that produced at 55°C. is in the thermal stability at 90°C. The 55°C. preparation shows only a 6–10% reduction in activity after 1 hour, whereas the 35°C. preparations show a 90–92% reduction in 1 hour.

Aside from the deletion of the second sentence and minor changes in verb tense and wording, the foregoing abstract is an exact replica of the summary that appeared in the original paper.

Following is another abstract of the same paper:

Conditions were determined for the production and purification of alpha-amylase from a strain of *Bacillus coagulans* and *B. stearothermophilus* after growth at 35° and 55°. The method increased the enzyme activity 450–600-fold. The only significant difference in the preparations produced at 35° and 55° was the thermal stability at 90°. The 55° preparations showed only a 6–10% reduction in activity after 1 hr.; the 35° preparations showed a 90–92% reduction in 1 hr.

In the second abstract, there was only a moderate editing of the author summary and a few changes of words.

The third abstract, which follows, comes closest to being completely original. It leaves out certain specific details contained in the author summary, and it includes details on method which were not included in the author's summary.

Amylases were isolated from *Bacillus coagulans* and *Bacillus stearothermophilus* grown at 35° and at 55°. After dialysis the activity of all preparations was reduced, and the enzymes were reactivated by Cl^-. The temp, optima were: for the 35° enzymes, 45–55°; and for the 55° enzymes, 60–70°. Heating at 90° had little effect on the 55° enzymes, but inactivated the 35° enzymes comparatively rapidly.

Further indication of the use to which author abstracts or summaries are put by abstracting publications is obtained from an analysis of the treatment of another paper, entitled, "Determination of Sodium Content of Human Sweat by Radioactive Sodium 24," by Decker, Genkins, and Braunwald. This paper was also picked up by *Chemical Abstracts, Biological Abstracts,* and *British Abstracts of Medical Sciences.* In two cases, the abstracts were signed by abstractors other than the authors. In the third case, the abstract was signed by the author. Following is the authors' summary which appeared with the original paper:

A method for determination of sodium concentration of sweat from local areas is described. It consists of administering radioactive sodium to subjects, and then inducing a local flow by subdermal injections of mecholyl chloride. Sweat so produced is collected by the method of Dole, Stall and Schwartz, employing filter paper discs sealed in a collecting device. Sodium concentration is derived by determining the radioactivity of the filter paper, and its increment in weight after sweat collection.

Following is an abstract of the specimen paper that appeared in one of the abstracting publications:

^{24}Na is given to the subjects, and a local flow of sweat induced by subdermal injections of mecholyl chloride. The sweat produced is collected on weighed filter paper discs sealed in a collecting device. Na concentration is derived by determining the radioactivity of the filter paper, and its increment in weight after sweat collection.

The foregoing abstract is an excerpted and slightly edited version of the summary that appeared in the original paper. It leaves out the first sentence of the summary, as well as details given in the third sentence as to the origin of the method used. It includes nothing that was not included in the authors' summary, aside from the fact that the filter paper used in the experiment was weighed in advance.

Following is a second abstract of the same paper:

A method for determination of Na concentration of sweat from local areas consists of administering about 100 microcuries of Na^{24} intravenously the night before, then inducing a local flow by subdermal injections of mecholyl chloride. Sweat so produced is collected by the method of Dole *et al.,* employing filter-paper disks sealed in a collecting device. The Na concentration is calculated from the radioactivity of the filter paper and its increment in weight after sweat collection.

In the second abstract above, the first two sentences of the author summary were combined, and details as to the amount of Na^{24} administered and when it was administered are given. Aside from that, the abstract is essentially identical with the authors' summary.

With the exception of one added phrase, one deleted phrase, and the addition of a final sentence about the advantages of the experimental method used, the third abstract, which follows, is a word-for-word copy of the authors' summary.

A method for determination of Na concentration of sweat from local areas is described. It consists of administering radioactive Na to subjects and after equilibrium is obtained, inducing a local flow of sweat by subdermal injections of Mecholyl Chloride. Sweat so produced is collected with filter paper discs sealed in a collecting device. Na concentration is derived by determining the radioactivity of the filter paper and its increment in weight after sweat collection. The chief advantage of this method is its accuracy and its avoidance of the necessity for general body heating.

The two specimen papers whose abstracting treatment was discussed in the previous paragraphs were both in the field of biochemistry. However, their abstracting treatment is typical of that given the majority of the specimen papers in the present study, regardless of field.

The apparent dependence of abstractors upon author summaries in the papers they abstract, regardless of field, is illustrated by the treatment given a paper by Anderson and Hutchison which was published in the *Physical Review*. The paper was entitled, "Paramagnetic Resonance Absorption in Praseodymium Trichloride." Following is the author abstract that was published with the paper:

Paramagnetic resonance absorption in crystalline solutions of $PrCl_3$ in $LaCl_3$ has been investigated at a frequency of 2.32×10^{10} sec` 1 and at the boiling point of He. The apparent *g's*, when *H* is parallel and perpendicular to the *c* axis have been found to be 1.79 and 3.98 respectively. The results have been discussed in terms of the known crystalline structure and Stark splittings.

The paper in question was picked up by *Chemical Abstracts, Nuclear Science Abstracts,* and *Physics Abstracts*. Two of the abstracts were copies of the authors' summary and were marked to indicate this. The third abstract was signed by an abstractor other than the authors of the original paper. This abstract follows:

Paramagnetic resonance absorption in cryst. solns. of $PrCl_3$ in $LaCl_3$ was investigated at a frequency of 23,200 Mc./sec. and at the b.p. of He. The apparent *g's,* when *H* is parallel and perpendicular to the *c*-axis, were found to be 1.79 and 3.98, resp. The results are discussed in terms of the known cryst. structure and Stark splittings.

Aside from some simple word and verb tense changes and a conversion from cycles to megacycles, the foregoing abstract is identical with the authors' summary.

THE QUESTION OF SUBJECT SLANTING

As was noted earlier, the vast majority of the signed abstracts examined in the present study were modifications of the authors' summaries that appeared with the original papers. To the extent that one is able to generalize from the sample used in the present study, it would appear that there is little subject slanting in abstracts of papers having author summaries or abstracts.

Assuming that the findings in the present study are based on a representative sample, a paradox arises. From an examination of the published items that *do* get original abstracts, it develops that these are primarily letters to the editor, communications, and similar short pieces. The apparent reason that such items get original abstracts is that they generally do not have author abstracts or summaries. Thus, long, detailed papers get author abstracts or refinements of author abstracts in the abstracting publications, while letters to the editor and the like get original abstracts when they are abstracted.

In view of the apparent limitations on the slanting of abstracts of papers with author abstracts or summaries, it is evident that the best place to look, with any degree of optimism, for subject slanting is in the abstracts of papers without author abstracts or summaries. An analysis was therefore made of the abstracts of 55 such papers. The abstracts for each of the 55 papers were carefully compared with one another in an effort to detect any distinct differences in subject viewpoint.

As it developed, there were 17 cases in which only one of the abstracting publications abstracted a paper and the others merely gave it a reference citation. Among the remaining 38 papers, there were 11 clear cases of subject slanting and 27 in which no slanting was detectable.

The three abstracts that follow are indicative of the abstracting treatment given an item having no author summary or abstract. The item, a letter to the editor of the *Physical Review,* was written by Z.Bay, V.P.Henri, and F. McLernon. It is entitled, "Upper Limit for the Lifetimes of Excited States of Ni^{60}." It was abstracted in *Chemical Abstracts, Physics Abstracts,* and *Nuclear Science Abstracts*. Two of the abstracts were signed by abstractors other than the authors. One was signed "auth." An investigation of the meaning of this latter inscription, in view of the fact that there was no author abstract accompanying the letter to the editor, revealed that the abstracting publication in question had used the first sentence of the letter to make up its abstract. As it developed, one of the two other abstracts had been done in exactly the same way, although in this case the abstract was signed. The third abstract departed from the other two in that it gave brief detail on the method and apparatus used in the experiment described in the letter. Following are the three abstracts:

The upper limit for the lifetimes of the 2 excited states of Ni^{60} following the β-decay of Co^{60} is 10^{-11} sec. The E_2 transition energies are 1.17 and 1.33 m.e.v.

An upper limit of 10^{-11} second for the lifetimes of the two excited states of Ni^{60} following the beta decay of Co^{60} has been determined. The E_2 transition energies are 1.17 and 1.33 Mev.

An account is given of a short-resolving-time (10^{-10} sec) coincidence circuit using diphenylacetylene scintillation counters. A value of 10^{-11} sec as an upper limit has been set on these lifetimes.

A further illustration of the abstracting of a paper with no author summary or abstract is obtained from an analysis of two abstracts of a paper entitled, "Metabolic Interrelations Between Intrinsic Factor and Vitamin B_{12}. III. Vitamin B_{12} Absorption at Varied Intrinsic Factor Doses." These two abstracts are as follows:

Intestinal absorption of B_{12} is increased in pernicious anaemia, when the intake of intrinsic factor is increased, but the dose of B_{12} unchanged. The increase in absorption occurs up to an optimal point, after which an excess of intrinsic factor shows no enhancing effect, or may even decrease absorption of the vitamin.

Tests were made on patients with pernicious anemia, whereby the hepatic uptake method previously described (C.A.48, 11590c) was utilized. Increase in oral dose of intrinsic factor prepns. from hog stomach or active human gastric material, with a const, dose of vitamin B_{12}, up to a certain point increased the intestinal absorption of the vitamin. Further increase seemed to have no further enhancing effect or even decreased the absorption of vitamin B_{12}. This last effect might be due to an excessive binding of the vitamin to the nonabsorbed fraction of the intrinsic factor prepn.

The first of the two abstracts above is from *British Abstracts of Medical Sciences,* and the second is from *Chemical Abstracts*. From a comparison of the two, one gets the impression that *Chemical Abstracts* is more concerned with clinical details than is *British Abstracts of Medical Sciences*. This would seem to be an example of "reverse slanting."

From the statistical analysis of the slanting of abstracts of papers having no author summaries or abstracts, and from the foregoing comparisons of the actual abstracts of such items, there would appear to be some subject slanting, but of a very minor nature.

CONCLUSIONS

1. Where there is an author abstract, use is generally made of it in the preparation of abstracts by abstractors for the indexing and abstracting publications studied.
2. Because of the heavy dependence on author abstracts as bases for abstracts in indexing and abstracting publications, a paradox develops in which short papers and letters to the editor, which generally do not have author abstracts, have the greatest likelihood of receiving original abstracts, while longer, more detailed papers have the smallest likelihood of original abstracting, because they generally do have author abstracts.
3. Author abstracts appear to be acceptable as abstracts for indexing and abstracting publications.
4. Where author abstracts are used, there can obviously be little subject slanting.
5. There is little detectable subject slanting in cases where original abstracts *are* prepared.

ACKNOWLEDGMENT

The author wishes to express his indebtedness to Mr. Robert S.Meyer for his very significant contribution to the planning and execution of the study upon which the present paper is based.

APPENDIX I SAMPLE JOURNALS AND THEIR ABSTRACTING TREATMENT

Journal titles	*Sample papers*	*Papers abstracted*	*Multi-abstracted papers*
Académie des Sciences, Paris, Comptes rendus	67	54	13
Acta Chemica Scandinavica	10	10	0
Acta Physiologica Scandinavica	3	2	0
Akademiia Nauk S.S.S.R., Leningrad, Doklady	40	28	6
American Chemical Society, Journal	50	50	9
American Journal of Botany	3	3	1
American Journal of Mathematics	3	3	0
American Journal of Physiology	12	12	12
American Mathematical Society, Transactions	4	4	0
Analytical Chemistry	16	16	3
Annalen der Physik	3	3	2
Annals of Botany	3	3	1
Annals of Mathematics	3	3	0
Archives of Biochemistry and Biophysics	8	8	8
Biochemical Journal	9	9	6
Botanical Gazette	3	3	2
Chemical Society of London, Journal	15	15	0
Chemische Berichte	6	5	0
Faraday Society, London, Transactions	4	4	1
Geological Society of America, Bulletin	3	3	1
Helvetica Chimica Acta	11	10	0
Helvetica Physica Acta	2	2	2
Journal de Physique et le Radium	4	4	3
Journal of Applied Physics	7	7	4
Journal of Biological Chemistry	13	13	12
Journal of Chemical Physics	18	17	17
Journal of Economic Entomology	10	9	0
Journal of Experimental Zoology	2	2	0
Journal of General Physiology	3	3	3
Journal of Geology	2	2	0
Journal of Physiology	5	5	5
Justus Liebig's Annalen der Chemie	3	3	0
Mathematische Annalen	2	2	0
National Academy of Sciences (Washington), Proceedings	4	2	1
Nature (London)	30	24	11
Nuovo cimento	6	6	3
Pfluger's Archiv für die gesamte Physiologie	2	1	0
Philosophical Magazine	5	5	4
Physica	3	3	2
Physical Review	30	30	27
Physical Society of London, Proceedings	6	6	6
Plant Physiology	5	5	5
Review of Scientific Instruments	6	5 .	5
Reviews of Modern Physics	2	2	2
Royal Society of London, Proceedings	9	9	8
Science	15	13	8
Société chimique de France, Bulletin	8	8	0
Société de Biologie, Paris, Comptes rendus	19	15	0

Society for Experimental Biology and Medicine, Proceedings	13	13	13
Zeitschrift für Physik	3	3	1
Zeitschrift für physikalische Chemie	3	3	0
Totals	516	470	207
Percentages of total sample	100	91	40

APPENDIX II Multi-abstracted papers

Ajzenberg, F., and T.Lauritsen, Energy Levels of Light Nuclei, Part V, *Reviews of Modern Physics, 27,* 77–166 (1955).

Alekseevskii, N.E., *et al.,* Use of a Heterogeneous Magnetic Field to Increase the Resolving Power of the Mass-Spectrometer, *Doklady Akad. Nauk SSSR Leningrad, 100,* 229–32 (1955).

Allfrey, V.G., *et al.,* Some Observations on Protein Metabolism in Chromosomes of Non-Dividing Cells, *Journal of General Physiology, 38,* 415–24 (1955).

Anderson, John H., and Clyde A.Hutchison, Jr., Paramagnetic Resonance Absorption in Praseodymium Trichloride, *Physical Review, 97,* 76–8 (1955).

Andrillat, Henri, Sur la Mesure des Intensités Relatives des Raies de O III et de la Raie H dans les Spectres des Nébuleuses Planétaires, *Comptes rendus, Académie des Sciences , 240,* 491–93 (1955).

Avignon, Yvette, and Simone Desprez-Rebaud, Étude, au Moyen de l'Emulsion Sensible, du Libre Parcours Moyen d'Absorption de la Composante Nucléaire de Rayonnement Cosmique, *Comptes rendus, Académie des Sciences, 240,* 515–17 (1955).

Azbel, M.Ya., Theory of the Skin Effect in a Constant Magnetic Field, *Doklady Akad. Nauk SSSR Leningrad, 100,* 437–40 (1955).

Baker, B.R., *et al.,* Puromycin. Synthetic Studies. VII. Partial Synthesis of Amino Acid Analogs, *Journal of the American Chemical Society, 77,* 1–7 (1955).

Banks, Charles V., *et al.,* Magnetic Studies of Nickel (II) and Palladium (II) Complexes with Some *Vic*-Dioximes, *Journal of the American Chemical Society, 77,* 324–25 (1955).

Bay, Z., *et al.,* Upper Limit for the Lifetimes of Excited States of Ni^{60}, *Physical Review, 97,* 561–3 (1955).

Bean, C.P., and D.S.Rodbell, Kinetics of Magnetization in Some Square Loop Magnetic Tapes, *Journal of Applied Physics, 26,* 124–5 (1955).

Bélanger, Leonard F., Autoradiographic Visualization of Ca^{45} Intake by Normal and Pathological Cartilage *in vitro, Proceedings of the Society for Experimental Biology and Medicine, 88,* 150–2 (1955).

Bentley, P.J., Some Aspects of the Water Metabolism of an Australian Marsupial *Setonyx brachyurus, Journal of Physiology, 127,* 1–10 (1955).

Bergkvist, K.E., *et al.,* Neutron Deficient Isotopes of Pb and Tl. II. Mass Numbers 204, 202, 201, and 200, *Philosophical Magazine, 46,* 65–9 (1955).

Berry, F.J., *et al.,* The Early Development of Spherical Blast from a Particular Charge, *Proceedings of the Royal Society of London. A227,* 258–70 (1955).

Berry, W.K., *et al.,* The *in vitro* Protection of Cholinesterases against Some Organophosphorus Inhibitors, *Biochemical Journal, 59,* 1–5 (1955).

Bettelheim, F.R., Amino Terminal Group in Chymotrypsinogen, *Journal of Biological Chemistry, 212,* 235–9 (1955).

Binggeli, Edmond M., Mesures de l'Intensité des Neutrons de la Radiation Cosmique, dans l'Eau, à Grandes Profondeurs, *Helvetica Physica Acta, 28,* 3–23 (1955).

Blicke, F.F., and G.R.Toy, Antispasmodics. XX. Basic 1,3-Dioxolanes and 1,3-Dioxanes, *Journal of the American Chemical Society, 77,* 31–32 (1955).

Boyd, R.L.F., and D.Morris, A Radio-Frequency Probe for the Mass-Spectrometric Analysis of Ion Concentrations, *Proceedings of the Physical Society of London. A68,* 1–10 (1955).

Boyle, W.S., and P.Kisliuk, Departure from Paschen's Law of Breakdown in Gases, *Physical Review, 97,* 255–9 (1955).

Braunstein, Jerry, and William T.Simpson, Overlaps of Trial Functions for the Hydrogen Molecule. II. Covalent and Ionic Character of H_2, *Journal of Chemical Physics, 23,* 176–8 (1955).

Breckenridge, C.G., and H.E.Hoff, Transmedullary Stimulation of Central Respiratory Mechanism in Apnea, *American Journal of Physiology, 180,* 219–31 (1955).

Brinkman, H.C., and B.Peperzak, Approximate Solutions of the Thomas-Fermi Equation for Molecules. II, *Physica, 21,* 48–52 (1955).

Brown, Callaway, *et al.,* Magnetic Susceptibility of Liquid Ozone-Oxygen Mixtures, *Journal of Chemical Physics, 23,* 103–08 (1955).

Brown, F.B., *et al.,* The Vitamin B_{12} Group. Presence of 2-Methyl Purines in Factors A and H and Isolation of New Factors, *Biochemical Journal, 59,* 82–86 (1955).

Brown, Fielding, and Charles L.Gravel, Domain Rotation in Nickel Ferrite, *Physical Review, 97,* 55–9 (1955).

Brown, Frederick C., Temperature Dependence of Electron Mobility in AgCl, *Physical Review, 97,* 355–62 (1955).

Brown, J.W., and C.H.Wadleigh, Influence of Sodium Bicarbonate on the Growth and Chlorosis of Garden Beets, *Botanical Gazette, 116,* 201–09 (1955).

Bube, Richard H., Photoconductivity and Crystal Imperfections in Cadmium Sulfide Crystals. Part II. Determination of Characteristic Photoconductivity Quantities, *Journal of Chemical Physics, 23,* 18–25 (1955).

Cade, R., An Electrostatic Problem Involving a Non-Linear Fluid Dielectric, *Proceedings of the Physical Society of London, B68,* 1–9 (1955).

Campbell, L.Leon, Jr., Purification and Properties of an α-Amylase from Facultative Thermophilic Bacteria, *Archives of Biochemistry and Biophysics, 54,* 154–61 (1955).

Cerulus, Frans, The Contribution of Non-Static Forces to the Binding-Energy of the Deuteron, *Helvetica Physica Acta, 28,* 67–84 (1955).

Chackett, G.A., *et al.,* The Distribution of the Products of Heavy Ion Reactions with Aluminium, *Philosophical Magazine, 46,* 1–11 (1955).

Checkland, P.B., and H.W.Thompson, Vibration-Rotation Bands of Deuterium Cyanide, *Transactions of the Faraday Society, 51,* 1–8 (1955).

Chiba, Yasutane, Two Components in Crystalline Chlorophyll Lipoprotein, *Archives of Biochemistry and Biophysics, 54,* 83–92 (1955).

Chu, T.C., and Edith Ju-Hwa Chu, Paper Chromatography of Iron Complexes of Porphyrins, *Journal of Biological Chemistry, 212,* 1–7 (1955).

Cinnéide, R.O., A General Method of Aminomethylation, *Nature, 175,* 47 (1955).

Clark, Irwin, A Colorimetric Reaction for the Estimation of Cortisone, Hydrocortisone, Aldosterone, and Related Steroids, *Nature, 175,* 123–4 (1955).

Cohen, Carolyn, Optical Rotation and Polypeptide Chain Configuration in Proteins, *Nature, 175,* 129–30 (1955).

Conway, James, The Behaviour of the Blood Pressure in Normal and Hypertensive Rabbits in Response to L-Noradrenaline and to Ganglion Block by Hexa- or Pentamethonium, *Journal of Physiology, 127,* 69–80 (1955).

Cormack, D.V., *et al.,* Irradiation of Ferrous Ammonium Sulfate Solutions: Energy Absorption and Ionization Calculations for Cobalt-60 and Betatron Radiation, *Journal of Chemical Physics, 23,* 162–4 (1955).

Crawford, John D., *et al.,* Mechanism of Renal Tubular Phosphate Reabsorption and the Influence Thereon of Vitamin D in Completely Parathyroidectomized Rats, *American Journal of Physiology, 180,* 156–62 (1955).

Crismon, J.M., and Ruth L.Dryer, Factors Influencing Quantitative Assessment of Adrenaline Threshold of Small Vessels in the Rat Mesocecum, *American Journal of Physiology, 180,* 1–8 (1955).

Crussard, Jean, *et al.,* Observation d'un Méson Lourd du Type θ Chargé dans les Emulsions Exposées au Rayonnement Cosmique, *Comptes rendus, Académie des Sciences, 240,* 261–64 (1955).

Cüer, Pierre, *et al.,* Nouvelles Preuves Experimentales sur l'Existence de Sous-Structures Instantanées dans les Noyaux Légers de l'Emulsion Photographique vis-à-vis des Nucléons Energiques, *Comptes rendus, Académie des Sciences, 240,* 75–7 (1955).

Curran, George L., Metal Chelating Agents and Hepatic Cholesterol Synthesis, *Proceedings of the Society for Experimental Biology and Medicine, 88,* 101–3 (1955).

Currier, H.B., *et al.,* Plasmolytic Studies of Phloem, *American Journal of Botany, 42,* 68–81 (1955).

Daion, M.I., Determination of the Mass of Charged Particles by Scattering and by Residual Length in Plates Situated Inside the Wilson Chamber, *Doklady Akad. NaukSSSR Leningrad, 100,* 453–54 (1955).

Danos, M., erenkov Radiation from Extended Electron Beams, *Journal of Applied Physics, 26,* 2–7 (1955).

Decker, Barry R., *et al.,* Determination of Sodium Content of Human Sweat by Radioactive Sodium, *Proceedings of the Society for Experimental Biology and Medicine, 88,* 60–4 (1955).

d'Espagnat, Bernard, États Excités à Très Courte Vie Moyenne et Invariance de Charge, *Comptes rendus, Académie des Sciences, 240,* 164–66 (1955).

Di Salvo, Nicholas A., Drinking Responses to Intravenous Hypertonic Sodium Chloride Solutions Injected into Unrestrained Dogs, *American Journal of Physiology, 180,* 133–138 (1955).

Djerassi, Carl, *et al.,* Alkaloid Studies. V. Synthesis of 1-Isopropyl and 1-Isobutyl-

2-Methyl-6, 7-Dimethoxy-1,2,3,4-Tetrahydroisoquinoline, *Journal of the American Chemical Society, 77,* 484–85 (1955).

Dominguez, A.Gonzalez, Les Parties Fines des Intégrales de Riemann-Weyl et les Precedes de Regularisation, *Comptes rendus, Académie des Sciences, 240,* 499–500 (1955).

Droop, M.R., Carotenogenesis in *Haematococcus pluvialis, Nature, 175,* 42 (1955).

Durrum, E.L., and S.R.Gilford, Recording Integrating Photoelectric and Radioactive Scanner for Paper Electrophoresis and Chromatography, *Review of Scientific Instruments, 26,* 51–6 (1955).

Ennor, A.H., *et al.,* Specificity of Creating Phosphokinase, *Nature, 175,* 120 (1955).

Eränkö, O., Distribution of Adrenaline and Noradrenaline in the Adrenal Medulla, *Nature, 175,* 88–9 (1955).

Ernster, Lars, *et al.,* Reconstruction of Coupled Phosphorylation in the Succinic Oxidase System, *Nature, 175,* 168–9 (1955).

Erspamer, V., Observations on the Fate of Indolalkylamines in the Organism, *Journal of Physiology, 127,* 118–33 (1955).

Feng, I-Ming, Analysis of the Effect of Various Factors on Metal Transfer and Wear between Specimen Pairs of Same Metal and Same Shape. II. Effect of the Surrounding Atmosphere, *Journal of Applied Physics, 26,* 28–32 (1955).

Feurt, Seldon D., and Lauretta E.Fox, Report on Wax from Several Species of *Tillandsia* and from *Ananas comosus* (L.) Merr., *Science, 121,* 42 (1955).

Flom, D.G., Apparatus for Studying Friction and Sliding Electrical Contacts, *Review of Scientific Instruments, 26,* 1–4 (1955).

Florey, Howard, Mucin and the Protection of the Body, *Proceedings of the Royal Society of London, B143,* 147–58 (1955).

Fornaguera, R.Ortiz, On Some General Properties of Static Solutions of Schiff's Equation, *Nuovo cimento, 1,* 132–58 (1955).

Fournier, Paul, De l'Action Métabolique des Glucides de Structure, *Comptes rendus, Académie des Sciences, 240,* 115–17 (1955).

Fox, H.Munro, The Effect of Oxygen on the Concentration of Haem in Invertebrates, *Proceedings of the Royal Society of London, B143,* 203–14 (1955).

Francois, André-Charles, and Maurice Michel, Action de la Pénicilline et de l'Aureomycine sur les Propriétés Désaminantes de la Flore Intestinale du Porc, *Comptes rendus, Académie des Sciences, 240,* 124–26 (1955).

Freier, H.E., *et al.,* Determination of Carbon and Fluorine in Highly Fluorinated Substances, *Analytical Chemistry, 27,* 146–49 (1955).

Friedman, Seymour, and Louis N.Heynick, Circuit for the Determination of Contact Potentials and Electron Temperatures from Retarding Field Characteristics, *Review of Scientific Instruments, 26,* 17–19 (1955).

Friedman, Sydney M., *et al.,* Specificity of Depressor Effect of Pitressin, *American Journal of Physiology, 180,* 57–60 (1955).

Fukushima, David K., *et al.,* The Characterization of Four New Metabolites of Adrenocortical Hormones, *Journal of Biological Chemistry, 212,* 449–60 (1955).

Garb, Solomon, *et al.,* Species Differences in the Utilization of Glucose for Contractile Force by Isolated Perfused Hearts, *American Journal of Physiology, 180,* 103–106 (1955).

Gergely, J., *et al.,* Fragmentation of Myosin by Chymotrypsin, *Journal of Biological Chemistry, 212,* 165–77 (1955).

Gilbert, D.A., and Victor W.Cohen, Nuclear Hyperfine Structure Interaction in 3.1-hr. Cs^{134m}, *Physical Review, 97,* 243 (1955).

Goldberg, A, *et al.,* Experimentally Produced Porphyria in Animals. (With Notes on Histopathological Studies.), *Proceedings of the Royal Society of London. Series B, 143,* 257–79 (1955).

Goodman, Joseph, *et al.,* Effect of in vivo Inhibition of Cholinesterase on Potassium Diffusion from the Human Red Cell, *American Journal of Physiology, 180,* 118–120 (1955).

Green, H., and H.S.Anker, Kinetics of Amino Acid Incorporation into Serum Proteins, *Journal of General Physiology, 38,* 283–93 (1955).

Greene, Inez, and E.P.Hiatt, Renal Excretion of Nitrate and its Effect on Excretion of Sodium and Chloride, *American Journal of Physiology, 180,* 179–82 (1955).

Gross, E.P., Shape of Collision-Broadened Spectral Lines, *Physical Review, 97,* 395–403 (1955).

Gross, Jerome, *et al.,* Extraction of Collagen from Connective Tissue by Neutral Salt Solutions, *Proceedings of the National Academy of Sciences, 41,* 1–7 (1955).

Grossfeld, Henry, *et al.,* Differentiation of Fibroblasts in Tissue Culture, as Determined by Mucopolysaccharide Production, *Proceedings of the Society for Experimental Biology and Medicine, 88,* 31–5 (1955).

Gustavson, K.H., The Function of Hydroxyproline in Collagens, *Nature, 175,* 70–74 (1955).

Haeusler, Claude, Étalonnage de la Bande de l'Ammoniac a 1.98μ, *Comptes rendus, Académie des Sciences, 240,* 507–508 (1955).

Hall, W.C., and M.A.Kahn, Influence of Temperature and Daylength on Growth and Sugar Content of Sugar Cane Following Auxin Treatment of Cuttings, *Botanical Gazette, 116,* 274–281 (1955).

Harding, G.N., Absolute Calibration of Neutron Counters with Po-α-Be, *Proceedings of the Physical Society of London, A68,* 52–3 (1955).

Hardy, Sarah, and Frank W.Putnam, Proteins in Multiple Myeloma. IV. Interaction with Metabolic *Nitrogen, Journal of Biological Chemistry, 212,* 371–8 (1955).

Haskins, Francis A., Changes in the Activities of Several Enzymes During Germination and Seedling Development in Corn (*Zea mays* L.), *Plant Physiology, 30,* 74–8 (1955).

Hauffe, K., and H.G.Flint, Über die elektrische Leitfähigkeit von Cadmiumselenid, *Annalen der Physik, 15,* 141–47 (1955).

Heber, G., Zu Tomonagas Methode der Behandlung Mittelstark Gekoppelter Felder. Methodische Studien. II, *Annalen der Physik, 15,* 174–77 (1955).

Hirshon, Jack M., and George K.Fraenkel, Recording High-Sensitivity Paramagnetic Resonance Spectrometer, *Review of Scientific Instruments, 26,* 34–41 (1955).

Ingersoll, Everett H., *et al.,* Urinary Bladder Response to Unilateral Stimulation of Pelvic Nerves, *Proceedings of the Society for Experimental Biology and Medicine, 88,* 46–9 (1955).

Ivanenko, D., *et al.,* Notes on a Nuclear Model, *Doklady Akad. Nauk SSSR Leningrad, 100,* 37–40 (1955).

Jaffe, H., *et al.,* Gamma and X-Radiation in the Decay of Am^{241}, *Physical Review, 97,* 142–50 (1955).

Jaffe, H.H., Some Theorems Concerning the Mutual Polarizabilities of Atoms and Bonds in Conjugated Molecules, *Journal of Chemical Physics, 23,* 1–4 (1955).

Jain, S.C., and K.S.Krishnan, The Distribution of Temperature Along a Thin Rod Electrically Heated *in vacuo.* V. Time Lag, *Proceedings of the Royal Society of London, A227,* 141–54 (1955).

Jerzy Glass, George B., *et al.,* Metabolic Interrelations between Intrinsic Factor and Vitamin B_{12}. III. B_{12} Absorption at Varied Intrinsic Factor Doses, *Proceedings of the Society for Experimental Biology and Medicine, 88,* 1–5 (1955).

Jessup, Ralph S., *et al.,* The Heat of Formation of Tetrafluoromethane, *Journal of the American Chemical Society, 77,* 244–45 (1955).

Joham, Howard E., The Calcium and Potassium Nutrition of Cotton as Influenced by Sodium, *Plant Physiology, 30,* 4–10 (1955).

Johnston, D.F., A Theory of the Hall Effect in Graphite, *Proceedings of the Royal Society of London, A227,* 359–67 (1955).

Jones, D.Rudd, and Jenifer Wignall, Acetylation of Sulphanilamide in Plants, *Nature, 175,* 207–8 (1955).

Kalman, Sumner M., and Hans H.Ussing, Active Sodium Uptake by the Toad and Its Response to the Antidiuretic Hormone, *Journal of General Physiology, 38,* 361–70 (1955).

Karjala, S.A., The Partial Characterization of a Histamine Metabolite from Rat and Mouse Urine, *Journal of the American Chemical Society, 77,* 504–05 (1955).

Karmakar, G., and L.Zechmeister, On Some Dehydrogenation Products of α-Carotene, β-Carotene, and Cryptoxanthin, *Journal of the American Chemical Society, 77,* 55–60 (1955).

Katz, Joseph, *et al.,* Analytical Procedures Using Combined Combustion-Diffusion Vessel, *Analytical Chemistry, 27,* 155–56 (1955).

Keii, Tominaga, Fine Structure of Olefine Deuteration, *Journal of Chemical Physics, 23,* 210–11 (1955).

Keim, C.P., Production of Enriched Non-Radioactive Isotopes at Oak Ridge National Laboratory, *Nature, 175,* 98–101 (1955).

Keller, William E., Pressure-Volume Isotherms of He^4 below 4.2°K, *Physical Review, 97,* 1–8 (1955).

Kenten, R.H., The Oxidation of Indolyl-3-Acetic Acid by Waxpod Bean Root Sap and Peroxidase Systems, *Biochemical Journal, 59,* 110–121 (1955).

Kikkawa, H., *et al.,* Nature of Pigments Derived from Tyrosine and Tryptophan in Animals, *Science, 121,* 43–7 (1955).

Kirkwood, John G., *et al.,* Intrinsic Viscosity of Flexible Macromolecules: Corrected Tables, *Journal of Chemical Physics, 23,* 213–14 (1955).

Kissman, Henry M., *et al.,* Puromycin. Synthetic Studies XI. D-Ribofuranosyl Derivatives of 6-Dimethylaminopurine, *Journal of the American-Chemical Society, 77,* 18–24 (1955).

Koshiba, M., and M.F.Kaplon, Analysis of a High-Energy Cosmic-Ray Shower. I. Soft Component and Trident Process, *Physical Review, 97,* 193–200 (1955).

Kuper, C.G., On the Thermal Properties of Fröhlich's One-Dimensional Superconductor, *Proceedings of the Royal Society of London, A227,* 214–28 (1955).

Kurti, N., and J.McIntosh, Reversible Generation of Temperature Waves ("Second Sound") in Liquid Helium II, *Philosophical Magazine, 46,* 104–5 (1955).

Kusserow, Bert K., A Mechanical Heart-Lung Apparatus with Gas Dispersion-Centrifugal Aerator, *Proceedings of the Society for Experimental Biology and Medicine, 88,* 161–5 (1955).

Lacam, A., Variations de la Vitesse des Ultrasons dans la Methane en Fonction de la Temperature et de la Pression, *Journal de Physique et le Radium, 16,* 172–3 (1955).

Lance, G.N., and E.C.Deland, On the Differential Analyser Solution of the Water Bells Problem, *Proceedings of the Physical Society of London, B68,* 54–5 (1955).

Langevin, Michel, Etude du Schéma de Désintégration de Ba^{133}, *Comptes rendus, Académie des Sciences, 240,* 289–91 (1955).

Lee-Whiting, G.E., Spectrum of Bi^{210} (RaE) and the Coupling Constants of Scalar and Tensor Interactions in Decay, *Physical Review, 97,* 463–8 (1955).

Lehmann, Pierre, and Jacques Miller, Interaction Quadrupolaire dans Cd^{111}, *Comptes rendus, Académie des Sciences, 240,* 298–99 (1955).

Lipton, Murray M., and Alex J.Steigman, A Simplified Colorimetric Test for Poliomyelitis Virus and Antibody, *Proceedings of the Society for Experimental Biology and Medicine, 88,* 114–18 (1955).

Loeb, Leonard B., Corona Threshold and the Range of Velocities of Pulse Spread in Geiger Counters, *Physical Review, 97,* 275–7 (1955).

Lukesh, Joseph S., Neutron Damage to the Structure of Vitreous Silica, *Physical Review, 97,* 345–6 (1955).

Lyons, J.B., Geology of the Hanover Quadrangle, New Hampshire-Vermont, *Bulletin of the Geological Society of America, 66,* 105–146 (1955).

Marrian, G.F., and W.S.Bauld, The Isolation of Epioestriol from the Urine of Pregnant Women, *Biochemical Journal, 59,* 136–141 (1955).

Mason, Howard S., and E.W.Peterson, The Reaction of Quinones with Protamine and Nucleoprotamine: N-Terminal Proline, *Journal of Biological Chemistry, 212,* 485–93 (1955).

Mason, W.P., Ultrasonic Attenuation Due to Lattice-Electron Interaction in Normal Conducting Metals, *Physical Review, 97,* 557–8 (1955).

Maxson, D.R., *et al,* Electron-Neutrino Angular Correlation in the Beta Decay of $Neon^{19}$, *Physical Review, 97,* 109–16 (1955).

McDonell, William R., and Sheffield Gordon, Decomposition of Methyl Alcohol by Co^{60} Gamma Radiation, *Journal of Chemical Physics, 23,* 208 (1955).

McWeeny, R., The Valence-Bond Theory of Molecular Structure. III. Cyclobutadiene and Benzene, *Proceedings of the Royal Society of London, A227,* 288–312 (1955).

Meyers, W.M., and H.F.Deutsch, Immunological Studies of Fetuin, *Archives of Biochemistry and Biophysics, 54,* 38–44 (1955).

Milburn, Richard H., Statistical Theory of Multiple Meson Production, *Reviews of Modern Physics, 27,* 1–14 (1955).

Miller, Sidney I., *et al.,* Raman and Infrared Spectral Data and Tentative Assignments for *Cis*- and *Trans*-Diiodoethylene, *Journal of Chemical Physics, 23,* 44–48 (1955).

Minz, Bruno, and Antonio Chamorro, Sur une Action Cortico-Hypothalamique

Déclenchée par l'Adrenaline, *Comptes rendus, Académie des Sciences, 240,* 454–55 (1955).

Morinaga, H., Effects of Isotopic Spin Selection Rules on Photonuclear Yields, *Physical Review, 97,* 444–6 (1955).

Morpurgo, G., and B.F.Touschek, Remarks on Time Reversal, *Nuovo cimento, 1,* 201–4 (1955).

Mount, L.E., The Ventilation Flow-Resistance and Compliance of Rat Lungs, *Journal of Physiology, 127,* 157–67 (1955).

Nagata, T., and S.Uyeda, Interaction of Two Constituents in Ferromagnetic Materials Showing Reverse Thermo-Remanent Magnetism, *Nature, 175,* 35–6 (1955).

Nagle, D.E., Scattering of 119-Mev Pions by Deuterium, *Physical Review, 97,* 480–6 (1955).

Nailor, Richard, *et al.,* Modifications in the Hydroxamic Acid Method for the Estimation of the Esterified Fatty Acids in Small Amounts of Serum, *Archives of Biochemistry and Biophysics, 54,* 201–05 (1955).

Nance, James F., and Irwin B.Perlis, Auxin Effects on the Utilization of C^{14}-Labeled Acetate by Wheat Roots, *Science, 121,* 104–5 (1955).

Nicholas, J.F., Point Defects and the Release of Energy from Deformed Nickel, *Philosophical Magazine, 46,* 87–97 (1955).

Nussbaum, R.H., *et al.,* Gamma Rays in the Decay of ^{112}Pd and ^{112}Ag, *Physica, 21,* 77–8 (1955).

O'Dell, Boyd L., *et al.,* The Antithyrotoxic Factor: Its Solubilization and Relation to Intestinal Xanthine Oxidase, *Archives of Biochemistry and Biophysics, 54,* 232–39 (1955).

Patterson, C., *et al.,* Age of the Earth, *Science, 121,* 69–75 (1955).

Pechet, Maurice M., Method for the Chromatographic Separation of Very Polar Steroids, *Science, 121,* 39–40 (1955).

Penner, S.S., and E.K.Björnerud, Experimental Determination of Rotational Temperatures and Concentrations of OH in Flames from Emission Spectra, *Journal of Chemical Physics, 23,* 143–52 (1955).

Perry, W.F., and Helen F.Bowen, Incorporation of Acetate into Fatty Acid by Liver Slices of Adrenalectomized Rats, *American Journal of Physiology, 180,* 21–22 (1955).

Porter, Richard F., *et al.,* Mass Spectrometric Study of Gaseous Species in the Si-SiO_2 System, *Journal of Chemical Physics, 23,* 216–17 (1955).

Potier, Robert, Sur le Développement en Integrates de Fourier, des Fonctions d'Onde des Corpuscules a Spin, *Comptes rendus, Académie des Sciences, 240,* 281–83 (1955).

Potts, W.J., Jr., Low-Temperature Absorption Spectra of Selected Olefins in the Farther Ultraviolet Region, *Journal of Chemical Physics, 23,* 65–72 (1955).

Pratt, William W., Angular Distribution of Protons from the Reaction Ti^{48} (d,p) Ti^{49}, *Physical Review, 97,* 131–2 (1955).

Putley, E.H., Thermoelectric and Galvanomagnetic Effects in Lead Selenide and Telluride, *Proceedings of the Physical Society of London, B68,* 35–42 (1955).

Ring, S.O., and L.M.Litz, Excitation Function for Na^{22} from Deuterons on Aluminum, *Physical Review, 97,* 427–8 (1955).

Robbins, Jacob, *et al.,* Electrophoresis of the Thyroxine-Binding Protein of Serum at pH 4.5 *Journal of Biological Chemistry, 212,* 403–10 (1955).

Robinson, D., *et al.,* Studies in Detoxication. 60. The Metabolism of Alkylbenzenes. *iso*Propylbenzene (Cumene) and Derivatives of Hydratropic Acid, *Biochemical Journal, 59,* 153–59 (1955).

Roderick, H., *et al.,* Decay Scheme of the Mirror Nucleus P^{29} and Related Results, *Physical Review, 97,* 97–101 (1955).

Roe, Joseph H., The Determination of Sugar in Blood and Spinal Fluid with Anthrone Reagent, *Journal of Biological Chemistry, 212,* 335–43 (1955).

Rojansky, V., Quantum-Mechanical Operators, *Physical Review, 97,* 507 (1955).

Rosin, H., and A.Farah, Post-Stimulation Potentiation of Contractility in the Isolated Auricle of the Rabbit, *American Journal of Physiology, 180, 75*–82 (1955).

Rothstein, Jerome, Dislocation Model of Liquids, *Journal of Chemical Physics, 23,* 218–19 (1955).

Rothstein, Morton, and Leon L.Miller, The Metabolism of -Aminovaleric Acid- -C^{14}, *Archives of Biochemistry and Biophysics, 54,* 1–5 (1955).

Ruetschi, Paul, and Paul Delahay, Hydrogen Overvoltage and Electrode Material. A Theoretical *Analysis, Journal of Chemical Physics, 23,* 195–9 (1955).

Rutman, Julia, *et al.,* Studies on Protein Synthesis *in vitro.* V. The Effect of Diet and Fasting on the Incorporation of Methionine-S^{35} into Liver Protein, *Journal of Biological Chemistry, 212,* 95–103 (1955).

Sagers, Richard D., and Jay V.Beck, Exchange of Radioactive Bicarbonate with the Carboxyl Group of Pyruvate by *Clostridium acidiurici, Archives of Biochemistry and Biophysics, 54,* 249–50 (1955).

Sawyer, Charles H., Rhinencephalic Involvement in Pituitary Activation by Intraventricular Histamine in the Rabbit Under Nembutal Anesthesia, *American Journal of Physiology, 180,* 37–46 (1955).

Schreiber, W.F., Oscillographic Measurements of Probability Distribution, *Review of Scientific Instruments, 26,* 88 (1955).

Schrödinger, E., A Thermodynamic Relation Between Frequency-Shift and Broadening, *Nuovo cimento, 1,* 63–9 (1955).

Searcy, Alan W., and Robert D.Freeman, Measurement of the Molecular Weights of Vapors at High Temperature. II. The Vapor Pressure of Germanium and the Molecular Weight of Germanium Vapor, *Journal of Chemical Physics, 23,* 88–90 (1955).

Seely, Ben K., Detection of Certain Ions in 10`$^{10}$-to 10`^{15}-Gram Particles, *Analytical Chemistry, 27,* 93–95 (1955).

Shore, B., Heparin-Activated Enzyme System of Human Plasma, *Proceedings of the Society for Experimental Biology and Medicine, 88,* 73–5 (1955).

Shull, C.G., and M.K.Wilkinson, Neutron Diffraction Studies of the Magnetic Structure of Alloys of Transition Elements, *Physical Review, 97,* 304–10 (1955).

Silen, William, *et al.,* Effect of Antibacterial Agents on Ammonia Production within the Intestine, *Proceedings of the Society for Experimental Biology and Medicine, 88,* 138–40 (1955).

Sinyakov, E.V., and I.A.Izhak, Influence of Mechanical Pressure on the Dielectrical Properties of Seignettoceramics, *Doklady Akad. Nauk SSSR Leningrad, 100,* 243–46 (1955).

Skolnik, M.I., and H.R.Puckett, Jr., Relaxation Oscillations and Noise from Low-Current Arc Discharges, *Journal of Applied Physics, 26,* 74–9 (1955).

Snell, Arthur H., and Frances Pleasonton, Spectrometry of Recoils from Neutrino Emission in Argon-37, *Physical Review, 97,* 246–8 (1955).

Solimene, N., and B.P.Dailey, Structure and Barrier Height of Methyl Mercaptan from Microwave Data, *Journal of Chemical Physics, 23,* 124–9 (1955).

Soloway, Saul, and Perry Rosen, Estimation of Basicity with a Novel Thermo-chromic Indicator, *Science, 121,* 99–100 (1955).

Speirs, R.S., and U.Wenck, Local Inflammatory Response Produced by Solutions Sterilized with Asbestos Type Filters, *Proceedings of the Society for Experimental Biology and Medicine, 88,* 89–93 (1955).

Steinberg, Robert A., Effect of Boron Deficiency on Nicotine Formation in Tobacco, *Plant Physiology, 30,* 84–6 (1955).

Stevens, Carl M., *et al.,* Preparation of S-Allyl-DL-Homocysteine and Related Compounds and Tests of Their Growth Effects in *Rats, Journal of Biological Chemistry, 212,* 49–57 (1955).

Stites, Joseph G., Jr., *et al.,* Preparation of Actinium Metal, *Journal of the American Chemical Society, 77,* 237–240 (1955).

Stone, David, and Oscar Hechter, Studies on ACTH Action in Perfused Bovine Adrenals: Aspects of Progesterone as an Intermediary in Corticosteroidogenesis, *Archives of Biochemistry and Biophysics, 54,* 121–28 (1955).

Stumpf, P.K., Fat Metabolism in Higher Plants. III. Enzymatic Oxidation of Glycerol, *Plant Physiology, 30,* 55–8 (1955).

Sugiura, Kanematsu, and C.Chester Stock, IV. Effect of O-Diazoacetyl-L-Serine (Azaserine) on Growth of Various Mouse and Rat Tumors, *Proceedings of the Society for Experimental Biology and Medicine, 88,* 127–9 (1955).

Swim, R.T., and H.E.Rorschach, Jr., Flow of Helium II through Narrow Slits, *Physical Review, 97,* 25–32 (1955).

Swiss, Edward D., and Clifton F.Mountain, Correlation of Emetic and Hypotensive Doses of Veratrum in Pigeon, Dog and Human, *Proceedings of the Society for Experimental Biology and Medicine, 88,* 19–22 (1955).

Tamm, Igor, *et al.,* Ultracentrifugation Studies of a Urinary Mucoprotein Which Reacts with Various *Viruses, Journal of Biological Chemistry, 212,* 125–33 (1955).

Telkkä, Antti, and Kimmo K.Mustakallio, Proximal or Distal Mercurial Inhibition of Succinic Dehydrogenase in the Kidney Tubules of Rat, *Science, 121,* 146 (1955).

Theis, W.R., Struktur des Phasenraumes und ein Zusammenhang Zwischen Bahndrehimpulsverteilung und Dichteverteilung der Nukleonen im Atomkern, *Zeitschrift für Physik, 140,* 1–15 (1955).

Thomas, R.G., Collision Matrices for the Compound Nucleus, *Physical Review, 97,* 224–37 (1955).

Thomson, J.D., Reaction of the Diaphragm to Denervation, *American Journal of Physiology, 180,* 202–204 (1955).

Tobailem, Jacques, Mesures Précises de Périodes Radioactives, *Journal de Physique et le Radium, 16,* 48–53 (1955).

Varshavskii, Ya.M., and S.E.Vaisberg, Correlations in the Equilibrium Distribution of Deuterium in the Isotopic Exchange of Hydrogen, *Doklady Akad. Nauk SSSR Leningrad, 100,* 97–100 (1955).

Visconti, A., Applications à la Théorie Quantique des Champs d'une Solution

Exacte de Certaines Equations Linéaires Operatorielles, *Journal de Physique et le Radium, 16,* 1–15 (1955).

Viswanatha, T., *et al.,* The Inhibition of Trypsin. II. The Effect of Synthetic Anionic Detergents, *Journal of Biological Chemistry, 212,* 301–09 (1955).

Wagener, J.C.S., Reduction of Alkaline Earth Oxides by Methane, *Journal of Chemical Physics, 23,* 204 (1955).

Wasserburg, G.J., and R.J.Hayden, Age of Meteorites by the A^{40}-K^{40} Method, *Physical Review, 97,* 86–7 (1955).

Wawzonek, S., *et al.,* Epiphyseal Plate Lesions, Degenerative Arthritis, and Dissecting Aneurysm of the Aorta Produced by Aminonitriles, *Science, 121,* 63–5 (1955).

Weale, J.W., Alpha-Gamma Angular Correlations and Internal Conversion Measurements in ThCC` and ThC`D, *Proceedings of the Physical Society of London, A68,* 35–42 (1955).

Weidmann, Silvio, The Effect of the Cardiac Membrane Potential on the Rapid Availability of the Sodium-Carrying System, *Journal of Physiology, 127,* 213–24 (1955).

Weinstein, L.H., and W.R.Robbins, The Effect of Different Iron and Manganese Nutrient Levels on the Catalase and Cytochrome Oxidase Activities of Green and Albino Sunflower Leaf Tissues, *Plant Physiology, 30,* 27–32 (1955).

White, F.A., *et al.,* New Naturally Occurring Isotope of Tantalum, *Physical Review, 97,* 566–7 (1955).

Wilkinson, J.H., and A.J.Feetham, The Biological Action of Substances Related to Thyroxine. 10. The Deiodination of Thyroxine in Thyroidectomized Rats, *Biochemical Journal, 59,* 21–24 (1955).

Wood, R.K.S., Studies in the Physiology of Parasitism. XVIII. Pectric Enzymes Secreted by *Bacterium aroideae, Annals of Botany, 19,* 1–27 (1955).

The Importance of Peripheral Publications in the Documentation of Biology

MILDRED A.DOSS

Adequate indexing of a field of science must be determined by the peculiar and individual needs of its workers. Publications which professional indexers and abstractors consider to fall outside that field often contain essential information which cannot be disregarded. Biology serves as an outstanding example of a science in which there is a necessity not only for indexing books and periodicals covering various aspects completely within this area of knowledge, but also for the thorough examination and indexing in great depth of publications in related or borderline fields, in order that the subject matter may be adequately and thoroughly covered.

There are at the present time a number of indices, in addition to the one used as an example in this paper, which attempt a broad coverage of scientific literature in order to supply specialized kinds of information to scientists working in their respective subject-matter fields. A few of these indices are the Gray Herbarium Card Index at Harvard; the Index Kewensis Plantarum Phanerogamarum of the Royal Botanic Gardens, Kew, England; the Zoological Record and Neave's Nomenclator Zoologicus of the Zoological Society of London; Schulze's Nomenclator Animalium Generum et Sub-generum, published by the former Prussian Academy of Science; and the Fungus Catalogue of the Crops Protection Research Branch, and the Index maintained by the Entomology Research Division, both of the Agricultural Research Service, United States Department of Agriculture.

It is the purpose of this paper to illustrate the importance of including borderline fields in biological indexing, by demonstrating the importance of these fields for indexing the literature pertaining to a specialized branch of biology, namely, medical and veterinary zoology.

MILDRED A.DOSS Animal Parasite Laboratory, Animal Disease and Parasite Research Division, Agricultural Research Service, United States Department of Agriculture, Beltsville, Maryland.

Medical and veterinary zoology, a major area of biology, roughly corresponds to the field of animal parasitology, but its ramifications extend into related areas, such as plant and free-living nematology and vectors of diseases transmissible to man and other species of animals. The zoological groups include parasitic Protozoa, Trematoda, Cestoda, Nematoda, Acanthocephala, parasitic Arthropoda, and minor groups, together with their invertebrate and vertebrate hosts.

The problem of indexing in medical and veterinary zoology has at least been partially solved, from both the basic and applied standpoint, by the utilization of a highly complex and unique, yet practical, indexing scheme for the storage and retrieval of information in this specialty. The complete Index is divided into five sections: an Author Catalogue, a Parasite or Subject Catalogue, a Host Index, a Checklist of Specific and Sub-specific Names, and an Anthelmintic Catalogue. These catalogues are housed and maintained in the Animal Parasite Laboratory, Animal Disease and Parasite Research Division, Agricultural Research Service, United States Department of Agriculture, Beltsville, Maryland.

During the 65 years the Index-Catalogue has been in existence, a period which begins with the first recognition of medical and veterinary zoology as a separate branch of biology, approximately 22,000 titles have been listed in our Key to Serial Abbreviations. The average increase in the number of publications indexed has been about 500 annually since 1950.

An examination of these publications revealed that 350, or less than 2 percent, are standard journals on parasitology, helminthology, protozoology, or veterinary medicine, whereas more than 98 percent represent publications in peripheral fields.

Among the peripheral fields are anthropology, bacteriology, botany, chemistry, engineering, evolution, fisheries, forestry, heredity, hunting, malacology, mining, nursing, nutrition, oceanography, radiology, refrigeration, sanitation, sewage, speleology, viniculture, and virology. Publications listed under these subject headings have at one time or another included facts or phenomena of parasitological importance, which indicates the great breadth of biological interests of scientists in these fields as well as the vast scope of the field of parasitology itself.

Parasitological information is also found in many types of publications such as reports and publications of livestock sanitary boards, quarantine services, departments of fish and game, reindeer industry, marine laboratories, pearl oyster fisheries, zoological parks, conservation agencies, tea research institutes, scientific academies and institutes, chambers of commerce, and expeditions; diplomatic and consular reports; reports on the Arctic and the Antarctic; compendia;

symposia; monographs; and Festschrifts. All these publications have contributed items of scientific importance in this field of biology.

Parasitologists and workers in the biological sciences in general must have, in addition to the usual material given in standard bibliographies and indices, certain other detailed information which should be made available for ready reference. This information which often lies buried deep within the body of a paper is a necessity for all who use the Linnean system of nomenclature, who are concerned with the International Rules of Zoological Nomenclature, the International Rules of Botanical Nomenclature, and the International Bacteriological Code of Nomenclature, and who are interested in the identification, transmission, and eradication of disease organisms or in developing new and improved varieties of plants and animals. Some of these items of information are:

a. Original descriptions of all new orders, families, genera, and species of plants and animals. With the growing interest of the International Commission of Zoological Nomenclature in higher categories, the recording of the original names should be expanded to include all major taxa.
b. Changes in nomenclature. These are constantly being made in the form of emendations and new combinations of specific names. As biologists obtain more knowledge concerning their specialties, current classifications change and records of these changes must be made. These regroupings and changes are part of the growing pains of science which must be recorded.
c. The exact date of issue of a publication. The date of issue must be ascertained as it often differs from the publisher's imprint date or date given on the cover of a periodical. If the printer's imprint of a Russian publication states that it was “authorized for publication December 30, 1953,” it is apparent that it could not have been issued in that year and a later date must be supplied even though the publisher's imprint date is 1953. Volume 6(1–2) of the *Indian Journal of Helminthology* bears the imprint date 1954, but the date of issue is given as March 28, 1956. All new species described therein must, therefore, be dated 1956. Many publishers have excellent records concerning dates of issue of books and periodicals, and it is often found necessary to write to them for this information.
d. Designation of type genera and species. This information must be given in any taxonomic paper in order to validate a name according to the Rules of Nomenclature. The question as to whether a species is a monotype, a type by original designation, or a type by subsequent designation must often be answered. Some branches of biology require records of type hosts and type localities.

e. Synonymy as given by the author of a paper. This information is important, especially if additional synonymy is recorded.
f. A record of the specific names which have been used within a given genus. This information is required when an author is naming a new species; otherwise he faces the possibility of creating a homonym.
g. The hosts from which a parasitic organism has been reported. These hosts may be experimental, intermediate, final, or merely mechanical vectors, and it is essential to know which group is involved.
h. Location on or within the host. This factor may be essential in determining the identity of the parasite, its pathogenicity, or measures for its control.
i. The locality in which a plant or animal is found. This information is of great importance in determining geographical distribution.
j. Certain essential morphological characteristics. These are considered necessary by entomologists for proper identification of insects.
k. The drugs or methods reported to have been used in controlling harmful animals and plants. This information is important in controlling pests.
l. Collateral information on the biology, physiology, and ecology of organisms and on the pathology of diseases caused by them. This information is of increasing importance to biologists.

Now that the scope and kind of information required in compiling a comprehensive index in the field of medical and veterinary zoology has been described, I should like to give a few examples of such material which have been obtained from “peripheral publications.”

An article entitled “Gigantorhynchidae brasileiros,” by Lauro Travassos, Brazil's foremost parasitologist, was the object of years of search until in 1952 it was located in *Annaes do Primeiro Congresso Medico Paulista realisado em São Paulo de 3 a 10 de Dezembro de 1916,* with an imprint date of 1917. Here we have an example of a systematic paper containing new genera and species of Acanthocephala appearing in the proceedings of a medical congress. In this same publication was found another important paper of which there was no previous record. This paper by Cassio Miranda recorded the presence of a nematode in the liver of a fish.

Not only must papers appearing in full in reports of congresses receive special attention, but abstracts of proceedings must also be carefully examined and indexed. The proceedings of the Indian Science Congresses have proved especially troublesome from a nomenclatorial standpoint, because of the policy of reporting new genera and species in abstract form with adequate description to give them validity. In 1954 there appeared in the *Proceedings of the Fortieth Indian Science Congress,* held in Lucknow in 1953, a series of five papers by J. Dayal and S.P.Gupta describing new genera and species of trematodes with sufficient description to render them valid according to the Rules of Zoological Nomenclature. Most of these genera and species appeared with full descriptions in volume 5(1) of the *Indian Journal of Helminthology,* March, 1953, pp. 1–80, under the authorship of S.P.Gupta. However, since the latter publication was not issued until January 15, 1955, the descriptions appearing in abstract form have priority.

The original description of the human cestode *Raillietina cubensis* was thought to have appeared in the abstract of a paper by Pedro Kouri in the Third International Congress for Microbiology, New York, Sept. 2–9, 1939, p. 176. Later, however, during a cursory examination of Gradwohl's *Clinical Laboratory Methods and Diagnosis,* 2d edition, 1938, pp. 1380–1385, an earlier description was discovered on which the name is now based.

The Russian publication *Soiuzpushnina* would ordinarily be classed with trade or farm journals. In 1931 there appeared in the November 1 issue of this publication, pp. 45–46, the description of a new parasite of the fox and raccoon (*Physaloptera sibirica* nov. sp.). The senior author, A.M.Petrov, is Russia's outstanding authority on parasites of wildlife, and his articles ordinarily appear in what would be classed as scientific and technical publications.

Names of societies are often misleading, but if an arbitrary scheme of indexing were set up it is doubtful if the *Proceedings of the Leeds Philosophical and Literary Society* would be assigned to parasitology. Yet in volume 4 (4), December, 1945, pp. 251–258, of these proceedings there appeared an article by L.Lloyd on “The Demonstration of Nuclear Division in Nematoda.”

Local history would likewise be considered out of the field of medical and veterinary zoology, but in *Glamorgan County History,* volume 1, 1936, pp. 401– 412, there appeared an article by H.W.Thompson entitled “Species of Lymnaeidae Affected by Parasitic Trematodes in Glamorgan.”

Institutes for medical research ordinarily confine themselves to investigations in the realm of human medicine, but occasionally they are the source of unexpected items such as the discovery of the intermediate host of *Taenia regis,* a cestode of the cat family genera *Felis* or *Leo*. In the *Report of the South African Institute of Medical Research for 1956,* p. 58, we find that the intermediate host of this parasite has now been determined to be *Oryx gazella*.

In 1947 the International Commission on Zoological Nomenclature became concerned for the second time with the relative status of the names of the genus of trematode worms responsible for the world's number one human tropical disease, schistosomiasis. In 1922 under Opinion 77 of the Official List of Generic Names in Zoology, *Schistosoma* Weinland, 1858, had been established as the accepted name of this genus. At the time this case was first considered, the Commission was unaware of the existence of the name *Billharzia* [sic], which had been published in 1856 by Meckel von Hemsbach in a book entitled *Mikrogeologie* and based its ruling on the use of the name *Schistosoma* by Weinland in 1858. The final ruling of the Commission (1947) established the name *Schistosoma* by use of its plenary powers while at the same time admitting the priority of *Bilharzia* but suppressing it. This difficulty could have been avoided if someone had taken the trouble to examine and index Meckel von Hemsbach's *Mikrogeologie*. The book was not unknown when Opinion 77 was rendered since the author citation had been published in 1907 in the Author Section of the Index-Catalogue of Medical and Veterinary Zoology.

Since more than 98 percent of the information available in the literature in this field of biology would be overlooked if “peripheral publications,” as defined in this paper, were not thoroughly searched for information pertaining to medical and veterinary zoology, it is concluded that adequate indexing in this field requires the well-organized and thorough searching of all scientific literature having the remotest connection with the principal sciences included in it. Without such coverage the adequate provision of the detailed information required by the scientists working in this area of biology would be impossible. Furthermore, the indexing system described illustrates that any indexing schemes used in the various biological disciplines must be designed by and for the scientists working in these disciplines in accordance with their own peculiar requirements and must not be set up for the convenience of the indexers themselves.

Current Medical Literature: A Quantitative Survey of Articles and Journals

ESTELLE BRODMAN and SEYMOUR I.TAINE

The traditional view of scientific, especially medical, literature is that its volume is so large that it has become impossible to bring it under control by traditional methods. So far as can be ascertained from a reading of the reports on this problem, most previous estimates of the size of the literature have been made either entirely on an a priori basis or by extrapolation from incomplete data. In the latter category, especially, belong estimates based on partial counts of the number of journals published.

Years of experience spent in examining thousands of medical serials convinced us that the use of the journal title as the quantitative unit of the periodical literature was both erroneous and misleading. Since it is the individual article that contains the specific medical information sought, it seemed logical to use this unit as the basis for determining the true size and therefore the indexing load. The results of some trial projects confirmed this subjective impression and also indicated the feasibility of a larger investigation based on the count of periodical articles within the separate journal issues. The present study was undertaken to procure these data; its further aim was to analyze the information obtained, to compare the results with some other studies made on the traditional journal title basis, and, finally, to draw some conclusion regarding the significance of the findings.

The National Library of Medicine makes a planned, regular, and continuing attempt to learn of the existence of and to obtain all the medical serials published throughout the world. As a result of this policy the Library at the end of 1957 held more than 13,000 different substantive and non-substantive serial titles of reasonable currency. For purposes of this study, it was assumed that

ESTELLE BRODMAN and SEYMOUR I.TAINE National Library of Medicine, Washington, D.C.

the number of medical serials, particularly those of a substantive nature, not received at the National Library of Medicine was so small and was probably scattered so widely in terms of subject, language, and country of origin, it would not distort the results obtained. The field of medicine is broadly defined by the Library; it includes the ancillary fields of dentistry, nursing, hospital administration, pharmacy, homeopathy, and osteopathy.

METHODS

A modest, part time investigation was set up in the National Library of Medicine. Because of limitations of time, staff, and equipment, only certain data were tabulated and these were analyzed for the most part by hand methods. All current serials coming to the National Library of Medicine for a three-month period (March-May, 1957) were initially divided into two categories: those indexed by the *Current List of Medical Literature* and those not so indexed. Since the information about the journals in the *Current List* was already available for previous operational purposes, nothing further was done with this group of titles. Material not indexed by the *Current List* was next sorted into two further groups: those containing substantive articles ("indexable journals") and those containing merely news items, abstracts, statistics, and other non-substantive miscellany ("non-indexable journals").

The number of individual journal issues in each group was counted, after which the non-indexable journals were discarded from this study; the remaining journals, the indexable journals, were analyzed for the following information:

1. Number of articles contained in the journal.
2. Periodicity of the journal and its articles.
3. Language or languages of the journal and its articles.
4. Country of origin of the journal and its articles.
5. Subject or subjects covered by the journal and its articles. (Subject categories used were adapted from the first edition of *World Medical Periodicals*.)

The information thus obtained was added to equivalent information for the journals and articles indexed in the *Current List of Medical Literature* to obtain the total picture. The findings were then compared with those of two earlier, somewhat similar studies, that of the Welch Medical Library Indexing Project (1) and that of the second edition of *World Medical Periodicals* (2) wherever possible.

FINDINGS

A total of 31,423 articles appearing in 2089 journal titles was counted in the three-month period for an annual total of 125,692 articles not indexed in the *Current List of Medical Literature*. By utilizing the periodicity data and by making an adjustment for the journals issued annually, semiannually, or at irregular intervals which would normally not be received in any single quarter, the annual number of journal titles and issues was estimated to be 2506 and 19,007 respectively. However, because all the current journals, irrespective of their subject fields, which were received in the National Library of Medicine during the three-month period were counted, many items are included which would not be indexed in a general medical index, which, by its subject scope, would be more restrictive than the collecting policy of the Library. For example, while material on bibliography, physics, and chemistry is admitted to the Library collection, it would be excluded from a medical index. If we remove from the count the out-of-scope items, the grand totals of indexable medical materials are reduced by 135 journals, 1165 issues, and 14,436 articles to the figures which appear in Table 1. Throughout the paper, we have used the maximum counts to be certain that we have not erred on the side of under-estimation. Furthermore, no deductions have been made for the considerable quantity of articles in journals in such non-clinical fields as general science, general biology, and psychology which would also not be indexed in a general medical index.

TABLE 1 National Library of Medicine Survey, 1957: Summary table

Substantive serial titles (per year)		
Number indexed in *Current List of Medical Literature*	1,508	
Number of additional titles found	2,371	
Total substantive serial titles		3,879
Substantive serial issues (per year)		
Number indexed in *Current List of Medical Literature*	11,434	
Number of additional issues found	17,843	
Total substantive serial issues		29,277
Substantive serial articles (per year)		
Number indexed in *Current List of Medical Literature*	107,478	
Number of additional articles found	111,256	
Total substantive serial articles		218,734

From these figures it appears that the total indexable medical periodical literature is just about double the size of the present *Current List of Medical Literature,* the largest medical index in the world.

PERIODICITY

The frequency with which the material is published is given in Table 2.

TABLE 2 Periodicity of medical literature

	Periodical Titles		*Periodical Issues*	*Periodical Articles*		*Periodical Articles per Title*
Frequency	*Number*	*%*	*Number*	*%*		
Weekly	64	1.6	3,588	18,354	8.3	270
Semimonthly	75	1.9	2,040	12,336	5.9	145
Monthly	1220	30.4	15,240	99,107	44.8	79
Bimonthly	639	15.9	3,994	33,122	14.9	51
Quarterly	953	23.7	3,892	29,975	13.5	31
Semiannually	126	3.1	252	2,149	1.0	17
Annually	369	9.2	369	6,303	2.8	17
Irregularly	568	14.1	1,136	20,280	9.1	36
Total	4014	99.9	30,441	221,626	100.3	—

It can be seen that in regard to productivity the rank order follows closely the frequency of appearance of the journal. The most frequently appearing type of journal (the weekly) contains more articles per year than the semimonthly, which in turn, contains more articles per year than the monthly, and so on down to the annual publication. This information, often guessed at and now confirmed, is of some interest. For example, the data indicate that whereas the total number of annual, semiannual, and irregularly appearing journals comprise more than a quarter of all journals by title, the indexing workload in terms of articles is actually under 13%. Stated another way, a decidedly misleading impression could be created by the isolated fact that a particular 1063 journals are indexed, since these yield a total of only 28,732 articles.

It is also interesting to note that while overall there is an average of 58.1 articles published per year in each medical journal, the titles now indexed in the *Current List* yield 71.3 articles per year per title. Also, the average number of articles per journal issue for *Current List* titles is 9.4 while the same figure for *non-Current List* journals is only 7.6 articles per issue. This would seem to indicate that *Current List* journals are more productive than *non-Current List* titles in numbers of articles they contain; if so, it would mean that indexing of the additional journals would not swell the total in direct proportions to the number added.

GEOGRAPHIC ORIGINS

During the three-month period, journals from 85 countries were received. The results are tabulated in Table 3. However, some of the smaller nations known to publish at least one medical periodical did not contribute any specimens during the collection period. Quantitatively, both by titles and articles, these omissions should not be significant.

TABLE 3 Geographic distribution of medical literature

	Periodical Titles			*Periodical Articles*		
In Current List	*All periodical titles*		*In Current List*	*All periodical articles*		
Country	*Number*	*%*	*Number*	*%*		
Algeria	3	6	★	291	327	★
Argentina	25	55	1.52	1,892	3,784	1.7
Australia	6	15	★	576	896	★
Austria	20	45	1.24	1,130	2,222	1.0
Albania	0	1	★	0	56	★
Belgium	36	52	1.44	1,791	2,691	1.2
Brazil	33	90	2.49	1,127	3,335	1.5
British West Indies	2	2	★	39	39	★
Bulgaria	5	5	★	387	387	★
Burma	1	1	★	29	29	★
Canada	20	55	1.38	1,494	2,662	1.2
Ceylon	1	2	★	15	37	★
China	6	14	★	422	998	★
Chile	6	14	★	270	642	★
Colombia	2	15	★	40	360	★
Costa Rica	1	2	★	15	47	★
Cuba	11	32	★	298	918	★
Canary Islands	0	1	★	0	24	★
Cyprus	1	1	★	10	10	★
Czechoslovakia	24	30	★	1,613	1,785	★
Denmark	19	41	1.14	1,076	1,824	★
Dominican Republic	1	2	★	12	92	★
Ecuador	1	7	★	16	176	★
Egypt	2	7	★	134	262	★
England	97	244	6.76	8,058	16,398	7.4
Finland	7	16	★	313	593	★
France	116	297	8.23	10,997	22,485	10.2
French Guiana	1	1	★	40	40	★
Formosa	0	1	★	0	48	★
Germany	136	315	8.72	12,449	25,741	11.7
Greece	2	11	★	65	449	★
Guatemala	1	3	★	40	96	★
Hawaii	1	1	★	36	36	★
Haiti	0	2	★	0	32	★
Honduras	1	1	★	20	20	★
Hungary	18	23	★	1,288	1,388	★
India	8	42	1.16	493	2,761	1.3
Indonesia	1	2	★	60	68	★
Italy	150	345	9.55	8,668	17,192	7.8
Iran	1	3	★	45	181	★
Israel	3	9	★	200	344	★
Ireland (Eire)	2	4	★	144	160	★
Iraq	1	2	★	20	36	★
Japan	22	187	5.18	1,081	19,665	8.9
Jordan	0	1	★	0	8	★
Kenya	1	1	★	66	66	★

	Periodical Titles			*Periodical Articles*		
In Current List	*All periodical titles*		*In Current List*	*All periodical articles*		
Country	*Number*	*%*	*Number*	*%*		
Lebanon	2	5	★	118	162	★
Luxembourg	1	1	★	5	5	★
Malaya	2	3	★	44	92	★
Martinique	1	1	★	6	6	★
Mexico	13	42	1.16	488	1,296	★
Morocco	1	2	★	180	180	★
Madagascar	0	1	★	0	72	★
Netherlands	28	63	1.75	1,733	3,017	1.4
New Zealand	2	9	★	74	230	★
Nicaragua	0	1	★	0	56	★
Northern Ireland	1	1	★	20	20	★
Norway	2	17	★	212	632	★
New Guinea	0	1	★	0	12	★
Pakistan	1	6	★	20	200	★
Panama	1	3	★	30	142	★
Paraguay	0	1	★	0	8	★
Peru	2	16	★	20	288	★
Philippines	3	13	★	130	450	★
Poland	34	44	1.22	1,756	2,396	1.1
Portugal	10	31	★	499	1,007	★
Puerto Rico	1	4	★	60	104	★
Rhodesia and Nyasaland Federation	1	1	★	48	48	★
Rumania	9	22	★	309	1,737	★
Saarland	1	1	★	12	12	★
Salvador	1	1	★	37	37	★
Scotland	4	6	★	134	150	★
South Africa	3	19	★	248	876	★
Spain	33	117	3.24	1,491	5,231	2.4
Sweden	17	42	1.16	1,755	2,567	1.2
Switzerland	59	112	3.10	2,863	5,203	2.4
Thailand	0	3	★	0	120	★
Trieste	1	1	★	31	31	★
Tunisia	1	1	★	90	90	★
Turkey	3	12	★	132	488	★
U.S.A.	391	847	23.46	31,751	53,975	24.5
U.S.S.R.	57	84	2.33	5,338	7,126	3.2
Uruguay	6	14	★	109	309	★
Venezuela	4	16	★	116	412	★
Yugoslavia	17	28	★	859	1,439	★
Total	1508	3597	—	107,478	221,626	—

★Under 1%.

By far the largest producer of medical periodical literature, on any basis, is the United States, which furnishes almost one-quarter of all the world's journals and/or articles. Although in number of journals Italy ranks second to the U.S. with 9.6%, with Germany (8.7%), France (8.2%), and England (6.8%) trailing in that order, on the basis of articles, Germany ranks second with

11.7%, followed in order by France (10.2%), Japan (8.9%), and Italy (7.8%). This qualitative and quantitative variation of the "Big 5" is made especially interesting by the leap into fourth place by Japan which now produces almost 20,000 articles yearly.

Oddly, a national pattern appears to emerge in regard to the average number of articles published in each journal title. As Table 4 indicates, the journals of certain countries are consistently characterized by a high number of articles, others are of a low productivity, and a third group can be described as medium producers:

TABLE 4 Average number of articles per title by country

High		*Medium*		*Low*	
Japan	105.2	India	65.7	Italy	49.8
U.S.S.R.	84.8	U.S.A.	62.6	Switzerland	46.5
Germany	81.7	Sweden	61.1	Spain	44.7

LANGUAGES

As shown in Table 5, twenty languages are utilized to convey practically all of the world's medical information in periodicals. The heading "Polylingual"

TABLE 5 Language breakdown of medical literature

	Periodical Titles				*Periodical Articles*			
In Current List	*All periodical titles*			*In Current List*	*All periodical articles*			
Language	*Number*	*%*	*Order*	*Number*	*%*	*Order*		
Chinese	4	12	.3	20	362	938	.4	15
Czechoslovakian	20	25	.7	14	1,470	1,690	.8	12
Danish	5	26	.7	13	368	924	.4	15
Dutch	10	33	.9	11	996	2,000	.9	11
English	607	1,375	38.2	1	45,651	82,687	37.3	1
Finnish	3	10	.3	21	134	402	.2	17
French	152	384	10.7	3	13,358	28,254	10.9	3
German	155	394	10.9	2	13,681	28,729	13.0	2
Hungarian	14	16	.4	18	1,037	1,137	.5	14
Italian	151	352	9.8	5	8,699	16,699	7.5	5
Japanese	1	128	3.5	7	108	17,232	7.8	4
Norwegian	1	13	.4	19	144	536	.2	17
Polish	31	39	1.1	10	1,702	2,302	1.0	10
Polylingual	110	132	3.7	6	5,900	7,316	3.3	7
Portuguese	42	121	3.4	8	1,584	4,356	1.5	9
Rumanian	7	19	.5	17	224	1,804	.8	12
Russian	56	75	2.1	9	5,290	6,930	3.1	8
Serbo-Croatian	14	24	.7	15	697	1,173	.5	14
Spanish	109	358	9.9	4	5,190	13,806	6.2	6
Swedish	2	22	.6	16	46	770	.3	16
Turkish	3	10	.3	22	132	456	.2	17
Other	11	28	.8	12	705	1,485	.7	13
Total	1508	3597	99.9		107,478	221,626	99.5	

refers to the journals which include articles written in more than a single language; "Other" includes all material in a single language other than the twenty already listed for which fewer than 10 journal titles were located.

Table 5 shows that English is by far the most common language (over 37% of the articles). Here again, as in the geographic breakdown, a similarly intriguing variation between journal and article counts emerges. In regard to journals, German ranks second with 10.9%, followed very closely by French (10.7%); Spanish and Italian run a close fourth and fifth, with 9.9% and 9.8% respectively. On the article basis, German is again second to English with 13%, just slightly ahead of French at 12.9%. Fourth place is, however, usurped by Japanese with 7.8%, and Italian is in fifth with 7.5%. Russian trails far behind in eighth place with a little over 3% of the total. Some quick arithmetic reveals the interesting fact that about 85% of the medical periodical literature is written in but six languages.

SUBJECTS

Any attempt to subdivide the field of medicine into its component parts is, at best, a frustrating procedure and rarely, if ever, does it produce a result that is completely satisfying. Much of the difficulty, apart from the ever-broadening scope of medicine, may be attributed to the inherent characteristics of the various related subject fields and specialties which resist a clean, sharp, and exclusive compartmentalization. For example, the subject field under which the greatest quantity of medical periodical literature falls is known as "general medicine." Actually, the specific articles which appear within these general medical journals can be distributed without strain among the different smaller subdivisions provided in the classification. Unfortunately, because of time limitations, the present study had to change its basis and restrict the subject breakdown of the collected material to the journal title and not, as in other cases, to the individual article itself.

Despite these solid reservations, the figures in Table 6 are still of interest and value, although more limited than might be hoped. Its greatest usefulness will probably be found in the data for the more easily circumscribed fields such as dermatology and ophthalmology.

COMPARISON WITH OTHER STUDIES

The findings of this investigation can be profitably compared with those of the Welch Medical Library Indexing Project and the second edition of the *World Medical Periodicals*. Unfortunately the latter work appeared just at the conclusion of the National Library of Medicine survey; it is to be regretted that

TABLE 6 Subject breakdown of medical literature

	Periodical Titles			*Periodical Articles*		
In Current List	*All periodical titles*		*In Current List*	*All periodical articles*		
Subject	*Number*	*%*	*Number*	*%*		
Alcoholism	2	12	★	55	227	★
Allergy	6	7	★	273	493	★
Anatomy (incl. embryology)	33	48	1.2	1,231	2,303	1.0
Anesthesiology	12	17	★	647	1,363	★
Anthropology	2	14	★	66	470	★
Antibiotics	8	11	★	635	687	★
Aviation Medicine	8	8	★	223	223	★
Athletics	0	10	★	0	524	★
Bibliography	1	20	★	52	596	★
Biochemistry	31	37	1.1	3,099	3,491	1.5
Biology, general	21	54	1.4	1,194	2,778	1.1
Beauty culture	0	2	★	0	56	★
Cancer	26	34	1.0	1,603	1,815	★
Cardiovascular system	29	43	1.1	1,426	2,614	1.1
Chemistry	1	37	1.0	12	9,260	4.1
Chiropody	1	9	★	60	268	★
Chiropractic	0	5	★	0	248	★
Chronic disease	1	1	★	108	108	★
Criminology	0	9	★	0	260	★
Civil Defense	0	1	★	0	64	★
Dentistry	4	180	4.5	456	8,296	3.4
Dermatology	29	50	1.3	2,117	3,625	1.5
Diabetes	1	7	★	51	187	★
Education	0	2	★	0	92	★
Endocrinology	17	21	★	1,192	1,396	★
Engineering	0	1	★	0	72	★
Enzymology	5	5	★	114	114	★
Experimental medicine	88	101	2.5	5,755	6,383	2.6
Food technology	1	10	★	49	669	★
Gastroenterology	18	29	★	1,167	2,779	1.1
General medicine	332	817	20.4	33,143	60,147	24.5
Genito-urinary system	19	27	★	1,103	1,743	★
Geriatrics	5	11	★	289	517	★
Gynecology and obstetrics	48	76	1.9	3,393	5,369	2.2
Hematology	17	24	★	786	1,322	★
Heredity and genetics	13	22	★	321	649	★
History of medicine	8	24	★	126	566	★
Homeopathy	2	22		208	972	★
Hospitals	8	47	1.2	957	2,949	1.2
Hydrology and climatology	1	14	★	24	764	★
Hygiene and public health	42	162	4.1	2,207	5,799	2.4
Hypnosis	0	2	★	0	44	★
Illustration, etc.	4	5	★	115	143	★
Immunology	11	11	★	421	421	★
Industrial hygiene	20	45	1.1	988	2,224	1.0
Infectious diseases	24	31	★	1,426	1,826	★

	Periodical Titles			*Periodical Articles*		
In Current List	*All periodical titles*		*In Current List*	*All periodical articles*		
Subject	*Number*	*%*	*Number*	*%*		
Internal medicine	14	21	★	1,091	1,631	★
Jurisprudence, medical	8	21	★	253	753	★
Leprosy	4	8	★	83	195	★
Malaria	3	5	★	80	140	★
Maternity and child welfare	0	14	★	0	532	★
Medical profession	3	13	★	177	953	★
Medical technology	10	19	★	435	755	★
Medicine and religion	0	8		0	360	★
Microbiology	42	54	1.4	2,728	3,360	1.4
Microscopy	4	7	★	133	313	★
Military and naval medicine	37	44	1.1	1,491	1,679	★
Mycology	1	2	★	2	150	★
Naturopathy	0	6	★	0	468	★
Neurology	50	71	1.8	2,625	3,433	1.4
Neurosurgery	11	13	★	454	642	★
Nursing	5	45	1.1	531	2,483	1.0
Nutrition	16	32	★	1,018	1,630	★
Occupational therapy	7	25	★	348	920	★
Ophthalmology	40	76	1.9	2,306	4,302	1.8
Orthopedics	18	44	1.1	1,022	2,194	★
Osteopathy	1	18	★	144	784	★
Otorhinolaryngology	35	60	1.5	2,146	3,250	1.3
Parasitology	16	23	★	748	1,288	★
Pathology	36	46	1.2	2,537	3,257	1.3
Pediatrics	59	92	2.3	3,778	5,694	2.3
Pharmacology	34	39	1.0	2,462	2,782	1.1
Pharmacy	15	113	2.8	767	4,487	1.8
Philosophy	0	2	★	0	76	★
Physics	12	26	★	1,328	2,956	1.2
Physiology	44	48	1.2	3,337	3,561	1.5
Physiotherapy	13	33	★	556	1,508	★
Plastic surgery	3	6	★	190	382	★
Population	0	3	★	0	180	★
Psychiatry	63	103	2.6	3,086	3,974	1.6
Psychoanalysis	8	15	★	292	468	★
Psychology	27	73	1.8	1,325	3,581	1.5
Plants	0	7	★	0	476	★
Radiodiagnosis, etc.	29	48	1.2	2,452	3,300	1.3
Red Cross	0	6	★	0	208	★
Rheumatism	13	20	★	535	659	★
Science, general	10	76	1.9	1,865	8,713	3.6
Serology	4	4	★	169	169	★
Sex	0	3	★	0	112	★
Social medicine	6	16	★	280	688	★
Sociology	0	13	★	0	920	★
Speech disorders	2	5	★	74	162	★
Statistics	0	2	★	0	264	★
Surgery	106	193	4.8	8,227	12,903	5.3

	Periodical Titles			*Periodical Articles*		
In Current List	*All periodical titles*		*In Current List*	*All periodical articles*		
Subject	*Number*	*%*	*Number*	*%*		
Therapeutics	10	49	1.2	624	2,768	1.1
Thorax	15	30	★	795	1,659	★
Tropical medicine	33	39	1.0	1,474	1,590	★
Tuberculosis	32	62	1.6	1,617	2,549	1.0
Veterinary medicine	5	47	1.2	532	3,732	1.5
Vitaminology	5	7	★	156	400	★
Zoology	1	14	★	90	1,146	★
Total	1809	3994	—	123,455	244,455	—

★Less than 1%.

time and staff were not available to do a thorough comparison of the two lists, especially in the language, country, and periodicity areas. It is to be hoped that further work can be done on this in the future. The Welch Library Indexing Project, sponsored by the predecessors of the present National Library of Medicine, carried on its work from 1948 to 1953. Among its other activities, the Project also undertook to do a comprehensive survey of the world's medical serials.

Table 7 gives comparative figures of the total number of journals studied by the three groups.

TABLE 7 Comparative counts of medical periodicals

Welch Medical Library Indexing Project Survey	4454
World Medical Periodicals	4360[a]
National Library of Medicine Survey	3879

[a] Current titles only.

The differences between the National Library of Medicine figure and those of the other two are not as significant as would appear at first glance. The Welch Medical Library figures are higher on account of differences in definition of current substantive periodicals. Actually, an adjustment on the basis of like criteria would probably effect an extremely close alignment; this is not surprising in view of the wide use made by the Indexing Project of the serial holdings of the National Library of Medicine.

The apparent discrepancy between the National Library of Medicine and *World Medical Periodicals* is almost entirely due to the coverage from two specific geographic areas in the National Library of Medicine. The National Library of Medicine is now overcoming a lag in the acquisition of materials from Latin-America and Japan occasioned by an earlier policy decision to collect selectively from these regions. Some 90% of the differential between the two totals appears to consist of Latin-American journals whose average output of articles is generally acknowledged to be quite low.

Tables 8 to 10 are comparisons of the results in terms of journal titles of the National Library of Medicine and of the Welch Medical Library Indexing Project. In spite of the time interval between the two investigations, there is a high degree of correlation; 0.905 for periodicity and 0.947 for languages, for example.

TABLE 8 Comparison of periodicity of medical journals

	National Library of Medicine Survey		*Welch Medical Library Indexing Project Survey*	
Frequency	*Number*	*%*	*Number*	*%*
Weekly	64	1.6	93	2.0
Semimonthly	75	1.9	95	2.0
Monthly	1220	30.4	1518	34.0
Bimonthly	639	15.9	603	14.0
Quarterly	953	23.7	868	19.0
Semiannually	126	3.1	73	1.6
Annually	369	9.2	385	9.0
Irregularly	568	14.1	822	18.4
Total	4014	99.9	4454	100.0

TABLE 9 Comparison of geographic origin of medical journals

	National Library of Medicine Survey		*Welch Medical Library Indexing Project Survey*	
Geographic Division	*Number*	*%*	*Number*	*%*
Europe	2015	55.9	2012	47.2
North America	944	26.2	1382	31.0
Asia	295	8.2	336	7.5
Latin America	283	7.9	523	11.7
Africa	38	1.1	45	1.0
Australasia	25	.7	39	.9
Other	0	0	27	.6
Total	3597	100.0	4454	99.9

TABLE 10 Comparison of languages of medical journals

	National Library of Medicine Survey		*Welch Medical Library Indexing Project Survey*	
Language	*Number*	*%*	*Number*	*%*
English	1375	38.2	2061	41.2
German	394	10.9	545	10.9
French	385	10.7	600	12.0
Spanish	358	9.9	608	12.1
Italian	352	9.8	348	6.9
Polylingual	132	3.7	0	0
Japanese	128	3.5	187	3.7
Portuguese	121	3.4	196	3.9
Russian	75	2.1	85	1.7
Polish	39	1.1	34	.7
Dutch	33	.9	49	1.0
Danish	26	.7	39	.8
Czechoslovakian	25	.7	38	.8
Serbo-Croatian	24	.7	22	.4
Swedish	22	.6	46	.9
Rumanian	19	.5	1	0

TABLE 10 Comparison of languages of medical journals

	National Library of Medicine Survey		*Welch Medical Library Indexing Project Survey*	
Language	*Number*	*%*	*Number*	*%*
Hungarian	16	.4	26	.5
Norwegian	13	.4	14	.3
Chinese	12	.3	1	0
Finnish	10	.3	17	.3
Turkish	10	.3	21	.4
Less than 10	28	.8	70	1.4
Total	3597	100.0	5008	99.9

SUMMARY

An investigation was undertaken to determine the approximate size and composition of present-day medical periodical literature by employing as the basic counting unit not the journal title, so frequently used in such investigations, but the journal article. Our object was to do away with the difficulty encountered in most previous studies, where certain assumptions had to be made about the relationship between journal titles and journal articles.

Data were collected on the number of journal titles and journal articles published in 1957, their periodicity and country of origin, the languages in which they were published, and their subject breakdown. Analysis of these data showed a direct linear relationship between the number of issues of a journal published (its periodicity) and the number of articles in it. On the other hand, there appears to be a true variation in the average number of articles published per journal title on the basis of the geographical origin of the journal, and linguistically, substantial deviations are encountered between journal title counts and article counts.

It was also possible to analyze the data of the medical periodical literature for the probable magnitude of what would logically be indexed in a general medical index. This turns out to be about 220,000 articles per year, or approximately twice the number already listed in the *Current List of Medical Literature,* the largest general medical index we now possess.

We believe the statistics presented here can be used meaningfully in additional ways, only a few of which are mentioned in this paper. It is to be hoped that further work can be undertaken in the future.

REFERENCES

1. WILLIAMINA A.HIMWICH and others. Survey of world medical serials and coverage by indexing and abstracting service, Welch Medical Library indexing project sponsored by the Armed Forces Medical Library. Baltimore, 1954.
2. LESLIE T.MORTON, comp. World medical periodicals, 2nd edition, World Medical Assoc., Geneva, 1957.

A Combined Indexing-Abstracting System

ISAAC D.WELT

The Cardiovascular Literature Project[1] was established in 1955 for the purpose of collecting, indexing, and disseminating detailed experimental and clinical information concerning the effects of chemical agents upon the anatomy, physiology and pathology of the cardiovascular system. The "raw" data are derived from the extensive available world literature.

The end result of this accumulation will consist of a series of handbooks in index format which are designed to act as a clear and authoritative guide to the world literature published during the last few decades.

The indexing procedures employed in this compilation are the result of a rather novel approach to the problems of storage and retrieval of scientific and technical information. They are based upon the assumption that existing methods do not meet the increasing needs of scientists for detailed, informative indexing. An attempt has been made to insure the retrieval of isolated "bits" of information of interest to the prospective user from among the complex, voluminous and polyglot literature of a specific area of medical science. In addition, such information is pinpointed with a great deal of precision so that the number of scientific papers which must be examined during the course of a search may be minimized.

ABSTRACTING

The conventional approach to the control of the scientific literature is almost invariably based upon the use of abstracts.

An abstract usually consists of a relatively brief and concise summary of the

ISAAC D.WELT Director, Cardiovascular Literature Project, Division of Medical Sciences, National Academy of Sciences—National Research Council, Washington, D.C.

[1] Supported by Research Grant H-2045 from the National Heart Institute, National Institutes of Health, U.S. Public Health Service.

contents of a scientific paper. It may be mainly of a descriptive nature, that is, it may present a general résumé of the main conclusions reached by the authors of the communication. It thus represents a detailed, expanded annotation. On the other hand, the abstract may be of an informative nature, presenting experimental data in some detail and quantitative information. The informative abstract is best represented in the Organic Chemistry section of *Chemical Abstracts*. In some cases, the abstract is so complete that an examination of the original paper by a searcher is almost superfluous. Such abstracts are usually compiled by subject matter specialists.

Within recent years, as a result of the explosive growth of the scientific literature, the informative abstract has assumed great importance. As it has become virtually impossible for the active scientist to read all of the original papers which are pertinent to his field of endeavor, he has begun to rely more heavily on secondary publications as an information source. The value of informative abstracting, carried on by individuals who are highly qualified scientists themselves, therefore becomes readily apparent.

Abstracts, however, also tend to become too numerous for efficient and regular perusal by the busy researcher. One of the complicating factors involves the classification of abstracts in the abstract journal. In many cases, multiple entries are indicated, since it is frequently impossible to classify satisfactorily an abstract of a paper dealing with a number of dissimilar subjects. Thus, for example, there is bound to be a difference of opinion concerning the placement of a particular *CA* abstract in Section 11F —Physiology rather than in 11G—Pathology or in 11H—Pharmacology. It seems to be largely a matter of the highly subjective decision of the editor in charge.

A number of scientific journals and abstracting periodicals make use of so-called authors' abstracts. In addition to summarizing their publication within the body of the article, authors are requested to provide an abstract of their contribution. Since most authors are not skilled in the art of abstracting, such abstracts are bound to be rather uneven in style as well as in content. They may constitute nothing more than brief annotations, several sentences in length. They may be purely descriptive, emphasizing what the authors *think* they have been successful in demonstrating. On the other hand, they may be very informative, indeed, but limited to a small segment of the publication itself. On the whole, they must necessarily be highly subjective. It is a normal human failing subconsciously to de-emphasize the significance of negative results or those not in accordance with theoretical preconceptions. Conversely, data assumed to be significant may be overemphasized. On the whole, the user of authors' abstracts cannot feel quite sure that the abstracts represent fair, undistorted résumés of the original publications. A meticulous literature searcher may fear that potentially important items may have been overlooked by the author, engrossed as he may have been in trying to establish his main thesis. This is particularly true in the numerous instances in which scattered, seemingly unimportant bits of data, accumulated as by-products of the main experiment, turn out to be of permanent importance to workers in seemingly unrelated fields.

In general, too much dependence on informative abstracts, even when written by impartial observers of the scientific scene, may lead to a false sense of security. In fact, in extreme cases, the lack of an abstract may prove to be advantageous. In order to illustrate this point, we may consider a searcher seeking for a highly specific piece of information, such as whether a certain chemical compound had ever been tested for a specific pharmacological action. He may find only one abstract reporting this compound and concerned solely with its purely chemical aspects. As a result, he may naturally assume that no biological tests had ever been performed and, as a result, neglects to look up the original paper. This scientific communication may have been abstracted by a chemist with a natural bias in favor of strictly chemical facts. He may have unwittingly overlooked a single sentence buried deep within the manuscript documenting the fact that the compound had indeed been tested biologically but had been found to be totally inactive. We thus have a case of irretrievable information due to a false sense of security on the part of the searcher. Many examples of significant data which have been similarly overlooked easily come to mind, resulting in completely unnecessary, time-consuming, and expensive duplication of effort.

As a result of the relatively high information content of the average scientific paper which must be compressed into the brief and concise format of the usual abstract, many subjective decisions must be made by abstractors concerning inclusion and exclusion problems. This, in the language of information engineers, is bound to generate a goodly amount of “noise.” Better abstracts would certainly result if each paper were abstracted separately by a number of abstractors and their efforts combined by a completely impartial editor. However, this is usually not the case as a result of economic factors and the necessity of keeping up with an ever increasing flow of papers to be abstracted. This state of affairs has led to the all too prevalent belief that any abstract is better than none at all. It has also given rise to the often-heard complaint that too many abstracting services are duplicating each other in their journal coverage. Such duplication is not only inevitable but is beneficial, up to a point. An abstracting service devoted to information of a chemical nature will provide abstracts from journals such as *Science* or *Nature,* which are markedly different from those compiled by a biological abstracting service.

INDEXING

Indexing constitutes, without doubt, the most important avenue of retrieval of scientific literature. Most abstracting journals provide comprehensive indexes to their abstracts, based upon the realization that without them, the utility of the abstracts would be greatly impaired.

The compilation of adequate indexes in any field of intellectual endeavor presents a number of highly complex problems to the documentalist. In the field of science, these difficulties are multiplied as a result of obscure and esoteric terminology, extensive subject specificity and, as mentioned above, the relatively high content of indexable information in comparison with the non-scientific literature. The indexer of scientific and technical publications cannot easily justify the omission of any bit of indexable information. There is always a good chance that someone may sometime wish to have this specific piece of data available to him.

"Word" indexing as opposed to "subject" indexing is of dubious value in the indexing of scientific periodicals. Subject indexing, on the other hand, must involve a more or less rigid standardization of semantic factors or terms. The lack of standardized subject headings, their undue flexibility or inconstancy from one year to another, is bound to lead to situations where irretrievability of sought-for information becomes a significant factor. The searcher's "threshold of frustration" can most easily be exceeded when he is obliged to search under a plethora of possible subject headings without the assurance that he will eventually find the proper one. A standard subject heading authority list should always be accompanied by as many "see" references as are necessary. Such references act as a glossary or dictionary and serve to direct the user to the proper heading from among a number of synonyms or "near synonyms." The differentiation between a synonym and a "near synonym" is a function of the analytical level of the index; that is, how specific and detailed are the subject headings themselves? It is believed that the number of subject headings should remain relatively constant as the scope of the indexing endeavor is diminished. In other words, if the area of information to be covered is decreased, those subject headings pertaining specifically to the new field of concentration must be further subdivided and refined, with the result that as many of them are used to describe the restricted area as were formerly employed for the less limited field of literature coverage. The construction of a suitable subject list is a task for the subject specialist only and cannot be based solely upon proper terminology. A compromise must be made with accepted usage in the interests of achieving easy accessibility. Since the prospective user must constantly be kept in mind, adequate provision of "see also" references must be assured. Such additional aids are based upon the "association of ideas" concept, and not necessarily upon obvious word relationships. The "see also" reference is a means of teaching the user how to use the index to his best advantage. He is given the all-important sense of security that he is quite likely to find the information he seeks even though he is not quite certain as to exactly what he is looking for when he first approaches the index. It is quite obvious, therefore, that the individual who sets up subject headings with their accompanying "see" and "see also" references must himself be a good representative of the prospective users of an index in terms of subject matter background and literature searching habits, practices, and approaches.

Many indexes, especially those accompanying abstract publications, are compiled from the abstracts alone. The indexer does not attempt to examine the original papers but relies largely upon the information contained in the abstracts. This leads to a rather unsatisfactory state of affairs. Any subject slanting, omissions, or errors on the part of the abstractor are perpetuated and magnified. The information provided by the index entry is now twice removed from reality, that is to say, the original paper. Additional "noise" generated by the indexer is added to the information system without the prior removal of that contributed by the abstractor. This practice of indexing abstracts is, of course, again traceable to the great mass of scientific publications which must be handled and to financial and personnel limitations. For the same reasons, index entries of trained, qualified indexers are rarely checked or duplicated by others with similar backgrounds and training. Index entries are usually rather telegraphic in style and are kept as brief as possible. As a result, here again, there may be a loss of information contained in the abstract which has under-gone indexing procedures. All that has been said about irretrievability resulting from abstracting holds true with respect to indexing and gives rise to an even more unsatisfactory state of affairs. All scientists must use an index as a means of preliminary scanning of pertinent literature. Therefore, when an indexer misses a potentially important item or indexes it inaccurately or obscurely, the item is rendered irretrievable for most intents and purposes.

When a relatively circumscribed area of science is indexed, the use of even the most detailed subject headings may still not be specific enough. The searcher may be directed to a dozen or so papers only to find that the information which he seeks is contained in no more than one or two of them. There is thus an avoidable loss of time on the part of the busy scientist and a certain dissatisfaction with the index. In addition to the use of standardized subject headings, therefore, other means of achieving the pinpointing of data must be used.

COMBINED INDEXING—ABSTRACTING

A possible solution to the problems discussed above can now be suggested. Essentially, it involves the combination of abstracts with index entries, thereby leading to the elimination of the conventional abstracts and the all too brief index entries. Instead, a detailed and relatively lengthy hybridized index entry results, containing much more information than the conventional entry and easily accessible by means of the alphabetized subject heading approach.

Conventional abstracts of an informative nature, containing all indexable items, can be relatively easily transformed into detailed index entries. The procedure involves the “dissection” of the abstract into a number of complete sentences, each standing alone as distinct “lines of data” and contributing a unique item of information. Each of these sentences can then be rearranged as to word order so that the standardized subject headings or subheadings may appear in a predetermined position. An example of this would be the following: “Compound A, when injected intravenously into anesthetized dogs, produces damage to their kidneys.” This sentence, extracted from within the body of an abstract, provides a useful item of information concerning the effects of Compound A in a specific situation and under special circumstances. The obvious subject headings would be “Compound A” and “kidney.” The above sentence is rewritten in indexable form as follows: “Compound A → Kidney, produces damage to, when injected intravenously into anesthetized dogs.” The arrow (→) which can also be used thus (←) is equivalent to “effect upon,” an item commonly used in conventional abstracts. Its direction is from the subject of the sentence to the object. The same information can also be written in a slightly modified form, as follows: “Kidney ← Compound A produces damage to, when injected intravenously into anesthetized dogs.”

As we have seen, accurate indexing can best be done by individuals who use the original paper rather than an abstract of it as their primary information source. It is therefore not much more time-consuming or expensive to index a piece of information according to the above method than to use the time-honored “effect upon” which would lead to the following conventional entries: “Compound A—kidney, effect upon, in dogs” and “Kidney=Compound A, effect upon, in dogs.” It can readily be seen that the additional information concerning the positive effects of Compound A; that is, that it really does produce damage, the importance of the route of administration (e.g., it may not produce damage when given by mouth) and the anesthetized state of the animal (which may alter the response to the administered compound as compared with a normal, unanesthetized dog) are available to the searcher. These data enable him to preselect the papers which he would like to consult in the original, to a relatively fine degree. The simple "effect upon" entry might refer to some twenty or so papers. Upon detailed inspection of these articles, however, it might be found that in only five of these was a positive effect of Compound A upon the dog kidney described. In the others, the proper experimental condition had not been present and the results therefore appeared negative or equivocal. The saving in time involved in pinpointing information and the lack of frustration cannot be overemphasized.

A binary (i.e., subject → object; object ← subject) indexing procedure logically gives rise to two index entries, for each "bit" of independent information. In our present undertaking, the subject is usually a chemical compound or group of compounds and the object, the biological entity, which may be an organ, an organ system, a physiological function, a disease or a symptom of a disease. The system of what we have come to call "reciprocal entries" enables the user to approach the indexed information either from the chemical side or from the biological point of view. Obviously, the binary or reciprocal system is not limited to such subject headings and can easily be utilized in the indexing of any cause and effect relationship or, for that matter, any two terms which are to be brought into coincidence. The provision of a second, reciprocal entry need not necessarily be carried out by the indexer himself. It is essentially an editorial job which can readily be done by clerical personnel. The use of standardized subject headings which can function either as subject headings in their own right or as subheadings, that is, can be placed either to the left or to the right of the arrow, facilitates the construction of a reciprocal entry by a clerical person since he or she can use a list of standardized headings.

Negative data can, of course, also be handled by this method. One need only add the additional terms "does not" or "did not" between the subheading and the active verb, to designate a negative result.

The extreme importance of recording negative data in an easily retrievable form has been stressed by many scientists. Unfortunately, however, even when such information is available in a published paper, relatively few abstracting or indexing publications make it a point to emphasize it or even to abstract or to index it. A good deal of lost motion, duplication of effort and the resulting expense and delay would be obviated if negative results were treated in the same manner as are the more popular positive reports.

Equivocal information is of value to the scientist insofar as it alerts him to the presence of "unfinished business" in the literature. Terms such as "increases somewhat" or "produces a slight degree of" can be used to index data which are still not well established.

The use of *active* verbs instead of passive ones must be encouraged in the interests of clarity and unambiguity. In the above example, if the entry were expressed as "Compound A → kidney damage produced by, upon intravenous administration in anesthetized dogs," it would lose much of its positive impact. The point of the arrow would have become, as it were, somewhat blunted. Our extensive experience with this system indicates that the use of active verbs does not in any way contribute a deterrent to the speed and productivity of the indexer. It certainly helps the user to understand the index entry more clearly.

It is possible and indeed, advisable, to construct a series of active verbs which may be used preferentially to describe the effects of the subject upon the object. General verbs such as "alters" and "affects" may be used where more detailed effects are not furnished by the author of a paper. Much more preferable is the use of verbs such as "increases," "decreases," "produces," etc., with their accompanying negative modifiers (i.e., "does not") whenever necessary.

Another advantage of such a list of preferred verbs is that it enables the sub-alphabetization of entries. In many instances, especially when a relatively restricted field of knowledge is under consideration, there is so much information available relative to a small number of subject headings that, no matter how detailed these headings may be, a large number of entries will accumulate under them. For example, in our own field of interest, there are hundreds of entries under the chemical, *epinephrine,* as a main heading and the physiological function, *blood pressure,* as a subheading. If the word sequence in the index entry is standardized so that the verb invariably follows the subheading, entries which are alphabetized first under the main heading and then under the subheading, can still further be alphabetized under the verb. If this practice is adhered to, a preferred list of verbs is necessary. Synonyms (i.e., increasesaugments) also must be considered and eliminated as far as possible. As a result, all entries concerned with an "increase" in the physiological function which is affected by the drug will be grouped together. Furthermore, all negative entries are under the "does not" heading with "alter," "decrease," "increase," and "inhibit" following in that order.

The provision of information concerning some of the conditions under which the specific effect was obtained is of help in further reducing the number of papers which must be examined by the searcher. The species of animal used, the route of administration, etc., are examples of these-special conditions.

The use of standard abbreviations serves to reduce the physical size of each index entry without at the same time cutting down its information content.

Admittedly, the average scientific paper will generate a large number of entries, if indexing is carried out in a detailed manner. On the other hand, however, abstracts will have been abolished without the irretrievable loss of their unique function, which is to summarize the paper as a whole. Each indexed paper is given an accession number, and each index entry resulting from the paper is referred back to the original by means of such a key number. It then becomes possible to assemble all entries referring to a specific paper either by manual methods (i.e., each entry having been entered on a separate file card) or by machine methods where a separate punch card is used for each index entry. Once the entries are assembled, they can easily be edited and an acceptable abstract prepared. Therefore, although the combined indexing-abstracting approach does involve "short-circuiting" the abstracting stage, abstracts can still be made available where necessary without the need for consulting the original paper.

This last point may be used to emphasize yet another unique function of the informative detailed index entry. Such a compilation may, to a limited degree, also function as a "handbook." A handbook may be considered as a terminal source of information, that is, its use eliminates the necessity of consulting the original paper. In many areas of science, particularly where quantitative information is of great importance (e.g., physical chemistry, toxicology), the index entry may furnish all the information necessary. Thus, for example, an entry such as, "Compound A—toxic effects, produces, in the dog at a dose of 500 mg. per kg., upon intravenous injection. LD_{50} is 700 mg. per kg. Symptoms are nausea, incoordination, etc," provides valuable information in its own right, without the necessity of consulting the original literature. The addition of the usual reference number may then be superfluous. This is particularly true of the many thousands of items of scientific information obtained as a result of preliminary "screening" tests, such as are used in research in the field of cancer chemotherapy and antibiotic control of infectious diseases.

The recording of negative data by this means would result in a great saving of time. Users of such an index would rarely find it necessary to consult the original paper, since reference to the index entry would assure them that the particular compound in question had indeed been tested in a certain animal species under specific conditions, by a particular route of administration, at one or more dosage levels and found to be inactive with respect to the biological response under study.

The combined "abstracting-indexing" approach to the control of the scientific literature is now being utilized successfully in the Cardiovascular Literature Project in the National Research Council. To date (May 1958) it has handled some 13,000 separate papers in numerous languages. These documents are largely reports of original laboratory and clinical findings but also include a number of reviews, monographs, and advanced textbooks.

In other words, our methods have been put to the test and have been found to be practical. Although much more research is obviously indicated, it is possible to handle a relatively large body of information by means of our present approach.

The retrieval question which is, of course, at the heart of the matter, cannot as yet be fully answered. As with most new approaches, the user must be taught how to take full advantage of the benefits inherent in the system. Its limitations must also be pointed out. If the proper questions are asked of the system, there is no doubt that it can function more effectively than existing abstracting and indexing services.

Instead of attempting to construct a means of literature control applicable to a wide range of subject matter, the present endeavor has been limited to a relatively well-defined area. This was done in order to establish the efficacy of this novel method in an area where the subject matter itself can most easily be handled by it.

On the other hand, however, there is good reason to believe that, with adequate modification, this approach could be extended to other areas of science and is not necessarily restricted to the chemical-biological sphere. Neither is it restricted to a binary system (i.e., chemical → biological; biological ← chemical). With relatively minor changes, three, four, or more items per "line of data" can be accommodated, each functioning, in turn, as main headings and subheadings. Thus a form of "coordinate indexing" can be achieved, with the important proviso that the index itself furnishes the necessary coordination instead of the user. A truly multidimensional index can therefore ultimately be achieved.

Although the present undertaking will result in a conventional printed index, it is easily amenable to machine methods. The coding problems have been minimized by the use of standard main headings, subheadings, and active verbs. Their translation into meaningful machine language is therefore a relatively simple problem. Mechanization of a combined indexing-abstracting system would be most helpful in facilitating the dissemination of the information which it can store, in addition to being an invaluable aid to its efficient retrieval. The publication of printed lists of index entries can be most easily achieved by the use of punch cards which accommodate both the typewritten entry, as well as its encoded version. Selection of desired entries can then be accomplished mechanically with a consequent saving of much time and effort.

In summary, the present contribution attempts to outline a new system for the control of the scientific literature which combines within itself the best features of conventional methods of abstracting and indexing. It has been demonstrated, at least to our satisfaction, to constitute a practical method for the storage of a relatively large volume of information concerning chemical effects upon biological entities. The use of this approach to facilitate the retrieval of indexed data appears, at this time, most promising. A great deal of effort has gone into attempts directed toward the mechanization of literature control within recent years. As a result, significant advances have been made, and numerous machine installations are now in operation. However, it seems to us that relatively little has been accomplished in the important area of abstracting and indexing which provide the raw material for machine processing. Mechanization ought to permit a much deeper and more detailed type of indexing than the manual approach without a proportional increase in costs. The savings effected by conversion to machines might well be used for the provision of more adequate indexing for, after all, no system of literature control, however intensively mechanized it may be, can ever be any better or more efficient than the accuracy of the raw material which it is called upon to process and the meticulous detail with which it is indexed for the purposes of storage and retrieval. This restores the "human factor" to the problem of the control of the ever increasing mass of scientific information, which is the basic problem of the present conference.

BIBLIOGRAPHY

1. WELT, ISAAC D. Indexing of Chemical-Biological Data in a Restricted Field of Medical Science. Presented at American Documentation Institute, 1955.
2. WELT, ISAAC D. Subject Indexing in a Restricted Field. *Science 123,* 723 (1956).
3. WELT, ISAAC D., and JUDITH T.MACMILLAN. The World Literature on Cardiovascular Drugs. *Bull. Med. Lib. Assoc. 46,* 60–72 (1958).
4. WELT, ISAAC D. The Detailed Indexing of Biological Effects of Chemical Substances. *Bull. Med. Lib. Assoc. 46,* July (1958).

A Unified Index to Science

EUGENE GARFIELD

The primary purpose of this paper is to discuss a plan for accomplishing what Neurath calls "an encyclopedic integration of scientific statements" (1), what I call a "Unified Index to Science." The intellectual values of such an index could be discussed at great length. However, the many social and philosophical implications and justifications for the "device" proposed here will be left to the sociologists, historians, and philosophers of science. (2) The justification for a unified index to science is made here strictly on its immediate practical value for improving scientific communication and documentation.

UNIFIED INDEX DEFINED

The term "Unified Index to Science" is intended to imply a single interdisciplinary index to *all* documents, primarily periodical literature in *all* fields of science. Interdisciplinary indexes are not new and revolutionary. Well-known examples are the *Industrial Arts Index* and *Die Bibliographie der Fremdsprachigen zeitschriften Literatur*.

NEW AND UNIQUE ASPECTS OF THE UNIFIED INDEX

A Unified Index to Science differs from existing indexes in scope. To be really effective, the Unified Index to Science must be as comprehensive as possible in coverage. It is in the borderland areas between specialties where such an index will not only provide a unique research tool, but also will fill a need that is not being satisfied by any of the current media of scientific communication. Since scientific research today is highly inter-disciplinary, the "selective" approaches of our traditional media, based on the old academic disciplines, can never give us anything more than makeshift tools, which do not function properly, considering the overall job to be done. Fragmentary approaches are not only inefficient but inadequate.

In addition to meeting the need for full and adequate coverage of scientific

EUGENE GARFIELD Eugene Garfield Associates, Philadelphia, Pennsylvania.

literature, the Unified Index should also include a comprehensive Citation Index (3,4). The Citation Index is not only a practical possibility, but also provides the reinforcement and extension required to complement traditional approaches to information. None of the existing indexing agencies provides such a service.

Despite the unique features of the Unified Index, it is anticipated that efforts of the various indexing services now extant would be utilized insofar as possible.

ADVANTAGES OF THE PROPOSED SYSTEM

The methodology suggested in this paper for compiling the Unified Index, regardless of its physical form, would offer the following significant advantages:

1. Provision for *one* logical *starting point* for all literature searches, regardless of subject.
2. *Standardization of nomenclature,* particularly in the areas of overlap between existing indexing services.
3. Provision for *detailed indexing* not possible in specialty indexes. An increased number of analytical entries per article would be economically and intellectually more feasible.
4. Elimination of all doubt as to whether individual articles had been indexed by specialty indexes, particularly in inter-disciplinary subjects where selectivity exercised by specialty indexes is necessarily arbitrary. *Complete coverage of all individual articles* becomes a practical possibility.
5. *Economic utilization of machines* for the compilation of the present specialty indexes and indexes to individual journals.
6. *Economic production and distribution* of scientific indexes by virtue of broadening the number of potential users. Mass production is the best known method for reducing product costs.

Numerous other by-products of the proposed plan, not to be discussed here, include production of detailed indexes to individual journals, such as cumulated indexes in English to individual foreign language journals, indexes classified according to whatever system is desired, polylingual indexes, and more detailed indexing for the existing indexes, both foreign and domestic.

CONVENTIONAL INDEXES TO BE USED FOR MANY YEARS TO COME

It is assumed that conventional printed indexes have a long life ahead, in spite of the current feeling that the recently published Decennial Index to *Chemical Abstracts* will be the last. By the use of modern techniques for printing and for storing information, such indexes can be continued and even expanded more economically. The innate limitations of such conventional a priori type indexes (5) should not blind us as to their present and continued usefulness for years to come. In combination with a posteriori indexing (5) exemplified by the Citation Index (3), conventional indexes, particularly if composited to form the Unified Index, will meet many day-to-day requirements of information retrieval. However, the Unified Index to Science provides one definite point of departure for more sophisticated scientific information processing.

THE UNIFIED INDEX DESCRIBED

A unified index to science could take many physical forms. In a large centralized science information center, this H.G.Wells type of "World Brain" might be a 3 by 5-inch card file, a random access electronic storage device, or a searching device such as Minicard or Filmorex. In this paper an alphabetical printed index is assumed. Subject headings (rubrics) like those found in the *Current List of Medical Literature* (CLML) would be followed by modifying sub-headings or "modifications" (6) such as those found in *Chemical Abstracts* (CA), or *Biological Abstracts* (BA). Headings would be followed by complete references such as found in the *Quarterly Cumulative Index Medicus* (QCIM), as well as a Citation Index (CI) reference. A typical entry might be:

ADRENALINE, antagonists to,

Brown, H.A., *J. Pharm. Pharmacol.* 42:1145–7 (1958), CI 367.

On page 367 of the Citation Index the same reference would be followed by the following:

J. Pharm. Pharmacol. 42:1145–7 (1958)

CA 42:994b, BA 35:4578, BrA 24:AIII, 601.

This means quite simply that the article has been abstracted by CA, BA and *British Abstracts* (BrA). The Citation Index would also provide a listing of all related bibliographical *descendants* of the article (5). The Citation Index has been discussed in previously published articles (3–5,15). However, this is the first time its use for consolidating references to and from the various abstracting services has been recommended. This feature of the Unified Index is significant. The significance of this feature can be illustrated by a concrete example. In Appendix I the differences between a CA and a BrA abstract for the same article are shown. The CA abstract is quite brief, whereas the abstract in BrA is more detailed. By the use of the CA subject index combined with the Citation

Index, the BrA abstract would be quickly identified. This could also be done by use of the author index to BrA, if you assume an abstract will be found there. However, it is not known which abstracting service will have covered the article in question until the author index *is* checked. Yearly author indexes may not be available for some time. This type of situation is even more important when the article is in a foreign language. Abstracts must sometimes be relied upon in order to decide on making a complete translation.

Let me stress, however, that a Citation Index is not necessary, obviously, to produce a composite index to science. The organization and preparation of a Unified Index emphasizes in a different way the values to be derived from a Citation Index as a supplement to the conventional indexes. Further, the Citation Index could be used to supplement the conventional media as they function in the dissemination of scientific information (5).

UNIFIED AUTHOR INDEX

An author index to the scientific periodical literature requires no great imagination either to visualize or to produce. It requires personnel with cataloging rather than scientific training. Linguistic difficulties in dealing with foreign names constitute the main problem of the author index compilation. The high value of a single author index to science should not be underestimated since an *author* is frequently the best "subject" approach one has to the literature.

COMPILING THE UNIFIED INDEX TO SCIENCE

In two published papers I have discussed specific techniques for the mechanical compilation of printed indexes and subject heading lists (7,8). In a paper specifically prepared for this International Conference, now attached as Appendix II, the same techniques are discussed in connection with the compilation of classified indexes from alphabetic indexing. By similar techniques made considerably simpler through technological developments during the past five years, the Unified Index to Science could be prepared economically and quickly. Modifications in the details of machine operations outlined previously would be necessary, but the same basic principles would be utilized. It is considered unnecessary to elaborate on these machine techniques except to mention that Hollerith punched cards could be easily combined with, replaced by, or supplemented by Flexowriters (tape perforating and tape operated typewriters), Listomatic cameras, Teletypesetters, Photons, etc.

COST OF THE UNIFIED INDEX

The high cost of preparing the index would be economically feasible. Such an index would have the widest possible market for a scientific indexing publication. The index would be used in the large inter-disciplinary libraries— sufficient justification for its compilation. It would also be used in the specialized scientific libraries, by individual laboratories, and if properly promoted, by individual scientists. Most institutions and/or scientists now subscribe to one or more of the various indexing services. At lower financial outlay each could now obtain a product of high quality and utility. Any publisher will testify to the “miracles” that can be performed if the sales volume justifies a low price.

PUBLISHING THE UNIFIED INDEX IN SEMI-MICROFORM

In order to satisfy the most skeptical critics, it is further proposed that the necessarily large Unified Index to Science could be *printed* by conventional printing methods, in a *semi-microform*. This would reduce the space and paper requirements by a factor of 16 to 1. It would more than offset the increased size of an index combining entries from a multiplicity of subject indexes. The high cost of paper, binding, and storage costs makes the use of micro-techniques imperative. Since indexes are reference tools for use in locating abstracts and articles to be read, the problem of readability is more easily resolved by simple enlarging techniques than is the enlargement of texts that must be carefully studied.

By the use of reduction ratios of 4 to 1, it is possible to reduce space and paper costs by over 90%. In addition, it is still possible for the user to employ inexpensive magnifiers which are commercially available in a wide range of types such as the readers used for enlargement of texts for the partially blind. At these reduction ratios, it is still possible to print by conventional methods, and it is also possible to read headings with the naked eye. Headings would be printed in a larger bold-face type. Once the desired heading were found, the 4X magnifier could be used to scan the individual index entries. This semi-micro method of printing scientific publications has great promise, I believe, for all types of large-scale publishing projects. We have been so obsessed with achieving high reduction ratios for microfilm storage that this simple intermediate approach has been completely overlooked, in spite of its successful use in wartime for publishing magazines.

SOURCE OF SUBJECT INDEX ENTRIES

The discussion above has been directed primarily at the problem of mechanical compilation and description of the Unified Index. One might assume that a centralized indexing service could prepare the Unified Index to the periodical literature. There are advantages and disadvantages in a completely centralized service. It is assumed here, however, that indexing and abstracting by the many individual indexing services would continue and even expand.

After an appropriate period of time the mechanical unit record for each article (punched card, file card, etc.) would be utilized to determine whether that article had been abstracted by at least one of the indexing services. It would be particularly interesting to study the overlaps, and gaps, in abstracting of articles in broad spectrum journals like *Nature, Science,* and *Endeavour.* Even the specialized journals are not always completely covered for one reason or another. Indeed, the editor of *Endeavour* has said (19):

It seems that, at present, about half the published articles are abstracted by the various abstracting services. The remedy might be a central indexing bureau, but its realization seems far away. Careful indexing by each journal of its own published articles, on some universally known classification system, would make the scrutiny of journals easier, and would help abstracting services in their selection of articles, enable them to extend the range of journals covered and contribute towards the reduction of the second time lag, namely that between the publication of a paper and the appearance of the abstract.

TREATMENT OF UNABSTRACTED ARTICLES

Unabstracted articles would then be assigned to an appropriate abstracting service. It is further assumed that each indexing service would modify its indexing procedures by indexing each article or abstract on a current basis, rather than after a cumulation of a year's abstracts. The method of indexing on a current basis is used by the *Current List of Medical Literature* (7). Recently a similar procedure was recommended to *Biological Abstracts* by this author (9). Information sheets containing all index entries for each article would be submitted to the indexing center. The mechanics of preparing indexing sheets have been discussed in detail (7,9). The indexing center would then mechanically sort and collate all index entries. Since each article and its associated index entries contain article identification, it is possible to combine all index entries for individual articles regardless of the source. In this way the different indexing approaches of each indexing service would be utilized to advantage for more detailed indexing. The indexing center would then compile the individual indexes to each specialty service, and subsequently would compile, as a by-product, the Unified Index, including the Citation Index, giving the complete abstracting record for each individual article. During the interval between publication of the article and appearance of the abstract as well as the index, there would be no gaps in the coverage of individual articles in the Unified Author Index and Citation Index. Unabstracted articles would be located indirectly by means of the Author Index and Citation Index. References for unabstracted articles would appear in the Citation Index listings for earlier articles previously abstracted and indexed. In this way the "reference chain" grows stronger, and few pertinent articles escape those who want to conduct a thorough search.

BENEFITS FOR PARTICIPATING ABSTRACTING SERVICES

The benefits derived for each of the participating abstracting services should be kept in mind. Men working in biochemistry would, through the Unified Index, use *Biological Abstracts* more effectively since the chemical indexing of CA would frequently turn up BA abstracts written from the biological, rather than the chemical point of view. CA abstracts would more often be referred to by biologists through BA indexing. In some searches, the biologist may be more concerned with the chemical aspects of the problem. Since abstracting services depend on the efforts of volunteer abstracters, the quality of abstracts must vary. *Redundancy* can be usefully exploited to make up for resulting deficiencies that creep in from individual preferences of abstracters.

IMPROVED NOMENCLATURE

A biological indexing service cannot afford to pursue the problems of chemical nomenclature as would a chemical index. A chemical indexing service does not pursue all the ramifications of biological nomenclature. By suitable cross-referencing, the Unified Index can give the user the best of both. Neither service has to concern itself any more than at present with nomenclatural problems. If it chooses, BA could obtain an index including chemical entries according to the CA system and CA could obtain the BA approach in biology. Where the twain will not meet, then repetition will probably prove the simplest way out. However, I see no reason why BA cannot use "Dibenzylamine" instead of "Dibenamine" as a main heading. Nor can I see any reason why CA must use "Adrenaline" when "Epinephrine" is equally useful. Standardization of terminology is a complicated problem. "The naive answer to this problem is that it only requires agreement" (10). However, this system offers an excellent opportunity for the meeting of minds because the problem is reduced to individual decisions about individual nomenclature situations rather than general principles. Authoritative indexes tend to produce increased standardizations as was exemplified by the decision of Eastman Kodak to use the CA system in its well-known catalog of 3500 compounds (11). Further, "stable" nomenclature (12) is an objective that is neither obtainable nor desirable. The proposed system, particularly if it includes a Citation Index helps to overcome the difficulties of a changing nomenclature (3).

LEGISLATION, GOVERNMENT, AND COMPLACENCY

In 1952 Avias (13) suggested legislation in order to bring about the resolution of scientific documentation problems. In the intervening years much progress has been made in the realization that governments must take an interest in these problems. The recent Congressional hearings are indicative of this. It would appear that those few heretics who have been crying in the wilderness for documentation centers capable of performing the tasks outlined in this paper may at last be vindicated (14).

THE GRANDIOSE SCHEME

Grandiose schemes always meet with excessive resistance, not because they are impossible to achieve, but because there are only a few with sufficient persistence to materialize their dreams and even fewer to carry them out. Ultimately, most large endeavors must fall by the wayside, to be replaced by others. However, their value at a particular stage of history cannot be disputed. We may have seen the last Decennial Index to CA. But who would dispute the value of the CA Decennial, Beilstein, Index Catalog, etc., now and for years to come!

A Unified Index to Science would be a convenience even if the strictly academic intra-disciplinary approaches were still the primary targets of today's research. In this era, however, of the biophysicist, the psychochemist, the human engineer, the instrumentation scientist, and the cosmobiologist, a Unified Index to Science is an absolutely necessary working tool for unfettered scientific progress.

REFERENCES

1. NEURATH, O. *International Encyclopedia of Unified Science,* Univ. Chicago, 1938. p. 20.

2. OPPENHEIMER, J.R. Reprint of interview given to E.R.Murrow, January 4, 1955. (Quoted in reference 5 below.)
3. GARFIELD, E. Citation Index for Science, *Science 122,* 108–11 (1955).
4. ADAIR, W.C. Citation Indexes for Science, *Am. Document. 6,* 31 (1955).
5. GARFIELD, E. Citation Indexes—New Paths to Scientific Knowledge. *Chem. Bull. 43*(4), 11`12 (1956).
6. BERNIER, C.L., and CRANE, E.J. Indexing Abstracts, *Ind. Eng. Chem. 40,* 725–30 (1948).
7. GARFIELD, E. Preparation of Printed Indexes by Automatic Punched-Card Techniques, *Am. Document. 6,* 68–76 (1955).
8. GARFIELD, E. Preparation of Subject Heading Lists by Automatic Punched-Card Techniques, *J. Document. 10,* 1–10 (1954).
9. GARFIELD, E. *Biological Abstracts* in an Era of Automation. Unpublished Report.
10. DYSON, G.MALCOLM. The Preservation and Availability of Chemical Knowledge *J. Chem. Ed. 29,* 239–243 (1952).
11. PATTERSON, A.M. Eastman Kodak Gets in Line, *Chem. Eng. News 30,* 4662 (1952).
12. MCATEE, W.L. Stable Nomenclature. *Science 116,* 400–1 (1952).
13. AVIAS, J. International Organization of Scientific Documentation Based on Legislation. *Science, 115,* 250 (1952).
14. GARFIELD, E., and R.L.HAYNE. "Needed—A National Science Intelligence & Documentation Center." Presented before the AAAS, Atlanta, Ga., December, 1955. (See *Chem. Eng. News 34,* 514 (1956).
15. GARFIELD, E. Breaking the Subject Index Barrier—A Citation Index for Chemical Patents. *J. Pat. Off. Soc. 39,* 583–95 (1957).
16. JOHNS HOPKINS UNIVERSITY. Welch Medical Library. Medical Indexing Project. Final Report, 1955.
17. GARFIELD, E. Preliminary Report on the Mechanical Analysis of Information by use of the 101 Electronic Statistical Machine. *Am. Document. 5,* 7–12 (1954).
18. GARFIELD, E. Preparation of Printed Indexes by Automatic Punched-Card Equipment—A Manual of Procedures. Johns Hopkins Univ. Medical Indexing Project. March 1953, 16 pp. (available from author).
19. The Accessibility of Knowledge, II. *Endeavour 11,* 4 (1952).

APPENDIX I COMPARISON OF THE INDEXING OF A SINGLE ARTICLE BY CHEMICAL ABSTRACTS AND BRITISH ABSTRACTS.

Chemical Abstracts 42, 6933f(1948).

Pharmacological and physiological aspects of adrenergic blockade with special reference to dibenamine. Mark Nickerson and Louis S.Goodman (Univ. of Utah, Salt Lake City). *Federation Proc. 7,* 397–409 (1948).

A review in the form of an address. 56 references. L.E.G.

British Abstracts, A III, 601(1949) same citation but no author address given.

Dibenamine (*N,N*-dibenzyl-2-chloroethylamine) (i) reverses adrenaline-induced blood pressure rise, (ii) reverses excitatory responses due to splanchnic nerve stimulation, (iii) reverses pressor effect of a generalized direct discharge of the sympathetic nervous system, (iv) blocks the response to reflexly elicited sympathetic discharge, (v) prevents cyclopropane-adrenaline arrhythmias, (vi) does not block or reverse inhibitory sympathetic functions, and (vii) only partially blocks the pressor effects of certain amines—e.g., amphetamine and Tuamin. Its chemistry, mechanism of action, therapeutic use, and relations to other drugs are discussed. B.G.Overell.

The article above was indexed by CA under:

Dibenzylamine, N-(2-chloroethyl)-, adrenergic blocking action of

The same article was indexed by BrA under:

diBenzylamine, *N*-ethyl-, 2-chloro-, in adrenergic blockade

Adrenaline, antagonists to

Adrenergic blockade,

In both indexing services there is a cross-reference from Dibenamine to the particular nomenclature preferred by each.

Experienced users of the indexing and abstracting services will not need to be convinced that there is considerable variation in indexing and abstracting from article to article. The example chosen here is not the best nor the worst. It was selected at random, to illustrate graphically several points.

1. Indexing under the same major subject heading by no means indicates that the indexing is the same or that an individual user will recognize the choices of the different indexers as being equivalent. Only a check of the actual article title would confirm that the article indexed was the same.
2. Abstracts vary in length, quality, uniformity, etc. In this instance the CA abstract is quite brief whereas the abstract in BrA is more detailed and useful.
3. Nomenclature varies in each case but in a Unified Index there would be no difficulty in making an arbitrary decision which to employ, if a choice must be made, without the loss of index access points. Both BrA and CA consider dibenamine a derivative of dibenzylamine, but the precise nomenclature is slightly different. (*N*-ethyl, 2-chloro- v. *N*-(2 chloroethyl). However, in BrA, all-benzylamines are grouped together whereas in CA they are not. From the user's point of view what is most important is locating the desired abstract in the quickest time. His first approach to the index will often depend upon his first encounter with the name. If it is a British article he is likely to be looking for the compound with British nomenclature. A Unified Index could eliminate redundancy where unnecessary and retain it where

useful, as in this instance. All cross references should be maintained, but it is obvious that in a composited index the number of total cross references for various subjects would be minimized by standardization.

4. The example also illustrates the different viewpoints brought out by different indexers. In BrA the article is indexed under "Adrenergic blockade" as well as "Adrenaline, antagonists to," whereas in CA it was not indexed at all under the subject of "Adrenaline." The indexing under "Dibenzylamine" was similar (adrenergic blocking action vs. in adrenergic blockade). Consequently in a combined index one of these two entries could be eliminated without loss to the user.
5. The example illustrates how a combined index could take advantage of different indexing, eliminate redunancy in indexing, reduce the searching time of the user, and standardize nomenclature for further ease of index use. It also demonstrates that neither abstracting nor indexing are ever precisely the same even in similar types of indexes, in this case two chemical indexes.

APPENDIX II MECHANICAL PREPARATION OF CLASSIFIED INDEXES FROM ALPHABETIC INDEXING

The pros and cons of alphabetic and classified indexes have been discussed quite extensively in the literature. It is not my intent to labor either view in this Appendix. During my work at the Johns Hopkins Indexing Project at the Welch Medical Library, it occurred to us that some useful compromise might be reached in dealing with the problem of alphabetic vs. classified indexes if one type of indexing could become the by-product of another. Project work on the preparation of subject heading lists clearly demonstrated the ease of converting alphabetic subject heading lists to categorized or classified lists of terms by the use of punched card equipment (8). In fact, the subject heading list of the *Current List of Medical Literature* used by the CLML staff is still maintained by this system.

Subsequently, the Project developed methods for the preparation of printed indexes by punched-card techniques (7). In this work the *Current List* was prepared *experimentally* by IBM machines, even though reports continue to appear which imply that the CLML is now prepared by machine methods. The project also did the pioneer work on the IBM 101 machine (17), and it was the combination of 101 capabilities and the system for preparing indexes by machine that led to the results reported here for the first time. The Project's support was terminated before all the data could be collected.

Figure 1 illustrates what I believe to be the first classified index to be prepared mechanically, by machine methods, completely as a by-product of alphabetic subject indexing. This is not mechanical classification; it is mechanical classifying in a limited sense only. Journal articles were originally read and indexed by *Current List* staff indexers by their present, i.e., conventional methods. This means that for each article, one or more subject headings were selected from the *Subject Heading Authority List* to describe the main subject matter of the article. For each alphabetic main heading, a standard sub-heading may or may not have been used. And for each entry a "modification" may or may not have been used (6).

Once this intellectual effort was completed, the necessary sets of punched cards were prepared as described in the Manual of Procedures (18). The cards were then mechanically sorted and edited, as described in the Manual, and the alphabetic subject index mechanically printed. To produce the classified index, the deck of punched cards, as well as cross-reference cards, were then mechanically sorted to produce the classified index illustrated in Fig. 1. The classified index was printed on an IBM 407 tabulating machine.

The sample shown is taken from category 2 (Anatomical terms), sub-division 2 (Organs), sub-division 6 (cardiovascular system), and includes subdivisions 1 (heart), 2 (arteries), 3 (veins). Thus, the complete class number for material on the heart is 2261. Under this classification are brought together all index entries for those subject headings which pertain to the heart. These include (in the sample) the cardiac septum, the heart itself, the mitral valve, the pericardium, pulmonary valves, and tricuspid valve. If the sample had been much larger, the grouping would include all terms which had been classified under this portion of the category list. The entries for each specific subject heading are exactly as they appear in the alphabetic index.

To accurately describe what has been accomplished, it should be noted that a method exists whereby the literature searcher can locate the same information provided to him by such a classified index. By use of the classified list of terms appearing in the *Subject Heading Authority List*—unfortunately the classified list is only available to *Current List* staff—he could determine all those subject headings pertaining to the heart and then locate these headings in the alphabetic subject index to the CLML. The classified index saves him the time of searching each of these terms individually. This is not intended as a criticism of the *Current List* which justifiably uses the most economical approach to serving the most people efficiently. Only with sufficient financial support and user demand can such classified indexes be justified. It is possible, however, that classified indexes would reduce the average number of subject headings required to index certain articles. The lack of indexing redundancy in the sample is a tribute to the excellence of CLML indexing efficiency. One might have expected that at least one or two articles would have been indexed twice under two closely related terms. This did not occur in the sample.

The mechanics of preparing the classified index from the alphabetical entries are quite simple. They involve a few wiring tricks with the sorting machines. It should also be noted that one could begin with a classified index and go over to an alphabetic index if desired.

During my work at the Project, I prepared a paper entitled "Unified International Scientific Indexes through Centralized Machine Indexing and Its Relation to Standardization of Nomenclature." Although this paper was approved by the then editor of *Science* and the referees, time was not available to complete the editorial revisions required. The complete paper was submitted to the present conference committee. Dr. Dwight E.Gray kindly suggested that the portion devoted to the problem of alphabetic vs. classified subject indexes be expanded for this Appendix. I am indebted to him for this suggestion and would like to close by summarizing the remarks I make on the subject of alphabetic and classified indexes in that paper.

The opinion is widely held that alphabetic indexes fail to group all related terms conveniently, i.e., they lack generic character. On the other hand, classification schemes rapidly outlive their usefulness, since they are unable to anticipate new subject fields. In practical terms, this means that the classification scheme must be

226 CARDIOVASCULAR SYSTEM
BLOOD VESSELS
GROWTH IN VITRO FR M BONE MARROW
CULTURE 24470
PRESERV FOR TRANSPL 24963
DISEASES
ENDANGIITIS OBLITE ANS RELATION TO
BUERGER S DIS 25308
INNERVATION
25216

2261 HEART
CARDIAC SEPTUM
INFARCTION
COMPL LEFT BUNDLE RANCH BLOCK ECG
DIAG 24464
RUPTURE
CAUSED BY NON PENETRATING CHEST INJ24461
HEART
EFF OF POSITION ON ECG 24465
EXTRACT EFF ON ISO ATED FROG HEART
VENTRICLE 25215
INVOLVEMENT IN PRO RESSIVE MUSC
DYSTROPHY 24451
INVOLVEMENT IN PRO RESSIVE MUSC
DYSTROPHY 24452
RHYTHM EFF OF POST RIOR PITUITARY
HORMONES DURING
CYCLOPROPANE ANESTH 24516
SIZE DURING HYPERTENSION 25231
ANATOMY AND HISTOLOGY
ATRIOVENTRICULAR N DE BUNDLE
CHANGES WITH ADVAN AGE 25462
BLOOD SUPPLY
CALCIFICATION 25218
X RAY 24426
DISEASES
AMYLOIDOSIS WITH CARDIAC FAILURE 25624
DIAG IN X RAY SURVEY FOR PULM TUBER24696
EFF OF AIR TRANSPO TATION OF
PATIENTS WITH 24688
HYPERTROPHY CAUSED BY GLYCOGENOSIS
X RAY DIAG 24603
IN PREGN 25585
MYOCARDIAL INVOLVEMENT IN SUBACUTE 24670
BACT ENDOCARDITIS PATHOL 24670
SURG ECG EPICARDIAL 25229
THROMBUS BALL OF AURICLE 24493
NEOPLASMS
LYMPHOSARCOMA CASE REPORT 24922
PATHOLOGY
POSTMORTEM EXAM IN GENERALIZED
SCLERODERMA 24641
PHYSIOLOGY
PRESSURE RECORDING IN DYING RABBIT 24458
MITRAL VALVE
HYDRAULIC FORMULA.FOR CROSS SECTION24453
AREA DURING REGURGITATION 24453
STENOSIS
EFF OF VALSALVA LI E MANEUVER ON
CIRC 25228
SURG INDIC - RESULTS 25184
VALVULOTOMY RESULTS COMPL 24643
PERICARDIUM
NEOPLASMS
TERTOMA IN NEWBORN 25519
PULMONARY VALVES
SURGERY
VALVULOTOMY AFTER CCLUSION OF PULM
ARTERY 24726
TRICUSPID VALVE
ABNORMALITIES
QUINTICUSPID 24395
2262 ARTERIES
AORTA
ABNORMALITIES
COARCTATION DIAG E AM OF ABDOM
AORTIC PULSE 24737
DISEASES
ANEURYSM ARTERIOSCLEROTIC SURG 24438
ANEURYSM OF HEPATIC ARTERY 24941
ANEURYSM SURG PRESERVED ARTERIAL
GRAFT 24450
ANEURYSMS - THROMBOSIS TECHNIC OF 24644
VEIN INLAY GRAFT IN THER 24644
ARTERIOSCLEROTIC A EURYSM THER BY
INTRASACCULAR WIRING 24637
EMBOLISM SURG 24942
EMBRYOLOGY
AORTIC ARCH DERIV IN ADULT 24630
ARTERIES
DISEASES
OBLITERATIVE DIS SURG THER 24639
SPASM PATHOGEN 25030
SPASM PATHOL - THER 25019
THROMBOSIS THER RESULTS OF SURG 24633
RADIOGRAPHY
PERIPHERAL ARTERIO RAPHY TECHNIC
DIAG VALUE 24638
ARTERIES CAROTID
EFFECT OF DRUGS ON
SYMPATOMIMETICS ON CAROTID SINUS
SYND 25207
ARTERIES FEMORAL
INNERVATION
MOTOR IN RABBIT 24956
ARTERIES ILIAC
DISEASES
ANEURYSM RUPT DURING PREGNANCY 24566
ARTERIES POPLITEAL
DISEASES
ANEURYSM SURG AORTA TRANSPL 24963
ARTERIES PULMONARY
DISEASES
ARTERIOSCLEROSIS C USING COR
PULMONALE 24598
STENOSIS
CASE HIST 25217
ARTERIES VERTEBRAL
DISEASES
SYND OF ANTERIOR SPINAL ARTERY 25612
RADIOGRAPHY
24884
2263 VEINS
VEINS FEMORAL
EFFECT OF DRUGS ON
ANTIBIOTICS ON BLOOD FLOW 25223
VEINS JUGULAR
DISEASES
PHLEBECTASIA DIAG PATHOL 24742
VEINS PORTAL SYSTEM
RADIOGRAPHY
24609
IODINE CONTRAST MEDIUM 24423
SURGERY
PORTACAVAL ANASTOM SIS IN PORTAL
HYPERTENSION 24725
VENAE CAVAE
SURGERY
LIGATION OF INFERI R EFF ON
OVULATION - PREGN 24547
PORTACAVAL ANASTOM SIS IN PORTAL
HYPERTENSION 24725
2264 LYMPHATIC
LYMPH NODES
DISEASES
GIANT FOLLICULAR H PERPLASIA IN
RHEUM ARTHRITIS 24763
LYMPHATIC VESSELS
RADIOGRAPHY
CONTRAST MEDIA PAT NT BLUE -
EVANS BLUE 25240
2265 HEMOPOEITIC
AND CHROMAFFIN SYSTEMS
SPLEEN
SURGERY
EXCIS FOLLOW UP 24484

FIGURE 1

revised periodically. These problems may be partially resolved by the system described above. This system evolved as a result of seeking a solution to the problem known in classification to some as "intercalation" or "interpolation" and by Ranganathan as "infinite hospitality." If one arbitrarily assigns a specific code number or some other notation to a concept, it is ultimately necessary to "intercalate" between two items in the schema more items than anticipated. This usually means considerable clerical work, if not a complete breakdown of the classification system. Through the "reproducing" features of some machines, it is possible to make such changes quickly. Special sorting techniques further simplify the procedure.

The preparation of a classified index of one type from an alphabetic list has been illustrated in this Appendix. However, it would also be possible to prepare other classified indexes from this same alphabetic index by the use of different class numbers or symbols in the master deck of subject heading cards. From the point of view of an international indexing center this capability could have considerable significance, since different classification schemes tend to bring out different approaches to subject matter. For example, a single deck of cards could produce one index according to the UDC system while another could be arranged according to the LC system. This capability could also be applied to classification systems such as chemical ciphering in which a single analysis of the chemical structure would enable codification according to numerous systems.

Contrary to the statements made in the original paper, I do not believe this system can handle, as easily as I had thought, changing terminology and concepts. I believe this problem is better resolved through Citation Indexing.

Lost Information: Unpublished Conference Papers

F.LIEBESNY

It is the practice of many learned and professional societies or institutions to publish the titles (and often short abstracts) of papers which were presented at meetings, symposia, conferences, congresses or conventions. On those occasions a large number of papers, often invited, is read. If, however, the abstract given in the reports of these meetings arouses sufficient interest in a reader to make him want to peruse a fuller account of such a paper, his chances are roughly even of being able to do so. The exact figure of 48.5% of unpublished papers has been arrived at by investigating 383 contributions to four conferences of American societies.

The papers presented at the following conventions were analysed: the 1948 and 1949 Annual Meetings of the Optical Society of America (called Meetings A and B hereafter), the 1949 National Convention of the Institute of Radio Engineers (Meeting C), and the meeting of the American Physical Society at Oak Ridge, Tennessee, on March 16–18, 1950 (Meeting D).

For this investigation the dates of the conventions were deliberately chosen as rather early so as to allow sufficient time for the papers to be published in full. In each case, the indexes to the journals in which the programmes appeared were consulted, and also the relevant abstract journals: *Chemical Abstracts and Science Abstracts A* for Meetings A and B; *Science Abstracts B* and *Wireless Engineer* Abstracts for Meeting C, and *Science Abstracts A* and *B* and *Chemical Abstracts* for Meeting D. Papers similar to the conference material or by fewer or more authors have been regarded as "identical" papers and counted as such.

It can be seen from Table 1 that about half of the papers read could not be traced as having been published in full. Of the published papers, about one third appeared in periodicals other than that in which the abstracts, of the papers appeared. It therefore seems necessary to scan more than one periodical. Furthermore, the time-factor is of great importance both to abstractors and

FELIX LIEBESNY The British Aluminium Company Ltd., London.

TABLE 1 *Analysis of the Papers*

	Presented	*Published by Dec., 1953*	*%*	*Published in other periodicals*	*%*	*Published[a] from date of programme publication* 1–12 months	13–24 months	25–36 months	More than 37
A 33rd Annual Meeting of Opt. Soc. Amer. (Oct., 1948)	76	34	45	7	20	22 (65)	4 (12)	6 (17)	2 (6)
B 34th Annual Meeting of Opt. Soc. Amer. (Oct., 1949)	62	36	58	6	17	27 (64)	4 (11)	3 (9)	2 (6)
C IRE National Convention (Mar., 1949)	172	88	51	33	27	42 (48)	41 (47)	3 (3)	2 (2)
D Amer. Phys. Soc. (Mar., 1950)	73	39	59	17	44	23 (59)	14 (36)	2 (5)	—
Total	383	197	51.5	63	32	114 (58)	63 (32)	14 (7)	6 (3)

[a] Figures in parentheses denote percentages.

readers. 90% of the papers eventually published were printed within two years of the appearance of the programme. Two of the 383 papers investigated were delayed by as much as four and a half years before they were printed in full. Sixteen of the published papers (8%) appeared in journals which are published by the firms employing the author (s) of the original paper, e.g., *RCA Review, Bell System Technical Journal, General Electric Review.*

To discourage any undue optimism in the potential reader, the list of summaries of the papers read at the Convention of the Institute of Radio Engineers is preceded by the following note: “No papers are available in preprint or reprint form, nor is there any assurance that any of them will be published in the *Proceedings of the I.R.E.,* although it is hoped that many of them will appear in these pages in subsequent issues”(1).

A similar warning prefaces the list of papers presented at the 1950 Winter Meeting of the American Institute of Electrical Engineers: “These papers are not scheduled for publication in *AIEE Transactions* or *AIEE Proceedings,* nor are they available from the Institute”(2).

Specially invited papers do not fare any better, as shown in Table 2.

TABLE 2 Analysis of invited papers

	Invited Papers		
Presented	*Published*[1]		
A	12	5	(42)
B	10	6	(60)
C	11	5	(45)
Total	33	16	(48.5)

[1] Figures in parentheses denote percentages.

Thus nearly half of the information presented at such meetings appears to be lost, unless some preprints are available, or unless the author is approached direct for copies of the manuscript.

This deplorable state of affairs is accentuated by the policies of the major abstracting journals; while *Chemical Abstracts* do present long lists of titles and authors, *Science Abstracts* do not publish abstracts "of papers presented orally at meetings, unless precis of substantial length are published in the normal way"(3).

The above analysis deals with American papers only, but there does not seem to be any great difference with affairs on the other side of the Atlantic. When papers are read at meetings of a British society, no abstracts are published in periodicals, but a small percentage of those papers (about 10%) is published in full in the proceedings of that society. Thus no appetites have been whetted, and therefore there are no disappointments. The Institute of Physics, however, publishes Conference Reports in its two journals. These reports are summarized proceedings of conferences organized by groups of the Institute; the reports are by one or two observers, but the actual authors of the papers are not indexed in the name indexes of the respective journals. The proceedings of some of the conferences are, however, published in full as Supplements to either journal (*Journal of Scientific Instruments* and *British Journal of Applied Physics*).

The importance of such "lost information" can perhaps best be illustrated by some examples:

At a meeting of the American Physical Society in 1926, J.C.Stearns read a paper about the effect of moisture on the viscosity of air (4). The six-line abstract published with those of the other papers presented at the meeting is the only source available for his findings of a decrease of viscosity with water vapour; this result is in direct contrast to the widely reported increase with moisture found by other scientists. For over thirty years it has been impossible to verify the accuracy of this statement. Since the author of this paper is no longer alive and the only known copy of the full paper is a typescript version in the Library of the University of Chicago, it seems unlikely that a serious enquirer will ever be able to make sure that there was not a two-letter misprint in the abstract, converting *in*crease to *de*crease. Yet for thirty-two years Stearns' controversial findings have been referred to and commented upon by writers who have apparently never seen the original paper—they all refer to the abstract!

Three scientists (one of them a Nobel Prize winner), M.McMillan, J.M. Peterson, and R.S.White, worked on the measurement of the ratio of negative *:* positive in the production of mesons; on theoretical grounds this ratio was expected to be unity, but had not been measured before. They published a result in the form of a pre-meeting abstract (5), which gave the ratio as approximately 10:1. At the meeting itself, the ratio had shrunk to 7.5:1 and was reported as such (6). By the time the complete investigation was written up, this ratio was down to 1.7:1 (6), i.e., reasonably close to the theoretical value. This example shows the potential unreliability of a hurriedly prepared pre-publication abstract, which may perhaps explain the above-mentioned controversial findings by Stearns, and also the danger of relying too much on such abstracts which yet may be the only information available.

The third example relates to the 132nd meeting of the American Chemical Society held in New York on September 8–13, 1957. This Society undertakes the publication of a booklet entitled: “Abstracts of Papers.” The 1957 meeting sported 1509 papers by 2700 authors and co-authors who alone would make a good-sized meeting. These papers were condensed into 680 pages in the above-mentioned booklet. The title of one paper, on “Russian Patents” aroused considerable interest; the abstract was barely five lines long and gave only general information about the subject. A direct approach to the author was made, which did not produce any reply. Yet this problem of obtaining full copies of U.S.S.R. patents which are now being abstracted in *Chemical Abstracts* still exists, and guidance from the author of that paper was not forthcoming, so that complicated and expensive arrangements had to be made to procure copies.

In the absence of preprints, which in any case were not available for general distribution prior to publication, any hope of obtaining further information was clouded by the prefatory statement that “presentation does not guarantee publication, but the majority of these meeting papers will doubtless be published. Many will appear in the publications of the Society.”

This statistical survey shows that the above majority is only just a mathematical majority by 1.5%, and that one third are likely to appear elsewhere than in the publications of the Society.

The state of affairs revealed by this analysis is rather alarming, when positive indication of the existence of information has been given, only to find that all the information consists of a title or a brief summary.

The question arises: what happens to the other 48.5% of the papers? Many are withdrawn completely or revised before publication. But the still considerable remainder leads one to the assumption that either the authors have died an untimely death (such as Mr. Stearns), or have suffered a severe attack of amnesia or laziness which prevent them from putting their knowledge into print.

UNESCO is also concerned with that problem, particularly the long delay which often intervenes between submission of papers for congresses and the printing. We live in an era of conferences and congresses, and yet allow so much of their results to be lost. Can we afford such a situation?

REFERENCES

1. *Proc. Inst. Radio Engrs., 37* (2), 160 (1949).
2. *Elec. Eng. (N.Y.), 69* (3), 250 (1950).
3. *Sci. Abstr. A58,* viii (1956).
4. *Phys. Rev., 27,* 116 (1926).
5. *Science, 109,* 438–439 (1949).
6. *Ibid., 110,* 579–583 (1949).

International Cooperation in Physics Abstracting

B.M.CROWTHER

In this paper a short survey will be made of the activities of the I.C.S.U. Abstracting Board (International Council of Scientific Unions) as seen from the point of view of *Physics Abstracts,* one of the two original member journals of the Board, with some comments on the effectiveness of these activities, and some suggestions for future development.

It will be useful first to recall the circumstances of the Board's foundation. At the Unesco Conference on Scientific Abstracting in Paris in 1949, a plan for a single universal international abstracting service had been rejected as impracticable at the present time. A more modest proposal, for the formation of a single international physics abstracting journal by the amalgamation of the only two journals dealing with this subject then existing was referred to a committee of the International Union of Physics. This committee reached the important conclusion that, however attractive might be the idea of a single international abstracting organization for physics, it was necessary, at the present time, for abstracting journals to be available in three or four different languages, to serve the interests of different language groups (English, Romance, Teutonic, etc.). It was recognised, in fact, that whatever facility the average scientist might have for reading original work in a foreign language, his reference book of compact information, with indexes in which the significance of an entry might often depend on a single word, had to be in a language very familiar to him. The committee was not able to recommend, at the present time, any practicable steps for the production and simultaneous publication in several languages of a single physics abstracting journal, though it expressed the hope that such a development might become possible in the future. It did, however, recommend the foundation and recognition by I.C.S.U. of a Board to assist the work of the existing physics abstracting organizations, both by encouraging and facilitating any joint action or mutual help which they might develop, and by its own independent activities, either independently conceived

B.M.CROWTHER Science Abstracts, The Institution of Electrical Engineers, London.

or suggested by the abstracting organizations. The Board would extend its recognition to only one physics abstracting journal in each language group, and thus try to preserve the relatively favourable circumstances at the time of its inception by discouraging the formation of directly competing journals.

The Board thus undertook the experiment, in a single subject field, of seeing to what extent an organic growth might develop, in a favourable atmosphere, between existing well-established institutions. Any attempt to impose a uniform pattern of behaviour on the member journals was explicitly excluded; it would have to grow, if at all, from the mutual agreement of the member journals, who were left to carry on their independent existence on the basis of whatever material resources they could command.

As will appear below, the degree of international cooperation which has developed is not sensational. In assessing its significance, it must be borne in mind that the Board had practically no financial backing, which set a strict limit on its executive activities; and that the abstracting organizations retained undiminished, throughout the period under consideration, all the major pre-occupations associated with the publication of an abstracting journal. In the case of *Physics Abstracts,* and, I have no doubt, the other organizations also, these were severe, arising primarily out of the rapid and continuing growth in volume and complexity of scientific publication. What has been achieved, therefore, has arisen from "spare time" work, and needs to be judged on that basis; the questions to be answered do not relate only to the degree of material advantage already gained, but also to the possibilities for the future, and to the requirements which experience has shown to be necessary.

The two founder member journals of the Board were: the *Bulletin Analytique* (later *Bulletin Signaletique*), published in Paris by the Service de Documentation of the Conseil National du Recherche Scientifique; and *Physics Abstracts,* published in London by The Institution of Electrical Engineers with the support of the British and American Physical Societies and the Institute of Physics. The *Bulletin Analytique* covered other sciences as well as physics, while *Physics Abstracts* is produced in conjunction with *Electrical Engineering Abstracts;* so that in each case the journals had other interests than physics alone. At the time of the foundation of the Board, these were the only two abstracting journals dealing with the whole field of physics. At a later stage, the German *Physikalische Berichte* reappeared, after the interruption of its publication in 1945, and this journal also was admitted to membership of the Board. The adherence of the Russian Institute of Scientific Information to the Board has occurred too recently for any significant developments to have taken place, save in the exchange of proof copies of periodicals. Other abstracting journals, not dealing with the abstracting of physics, who have been invited to join the Board, will not be considered in this survey.

Among the first acts after the establishment of the Board was to arrange for the member journals to exchange copies of their respective publications, and to grant general permission for the reprinting of abstracts. For reasons which will be discussed later, *Physics Abstracts* has not, up to now, made use of this. Also exchanged were lists of periodicals being regularly scanned for physics articles, and this brought to light a number of unfamiliar titles, about which we made enquiries, and also about other titles which had come to our notice through other channels; we received similar enquiries from the other member journals. These enquiries have continued from time to time, and it has been most valuable to be able to obtain a "professional" opinion on the status and contents of a periodical from someone engaged on the same business as oneself, who by proximity has better opportunities for becoming familiar with it and possibly establishing personal contacts with the editor. The standards of selection and the range of subjects covered are not necessarily identical in all respects for the member journals, so that in many cases it has in the end been necessary to form one's own judgement by inspection of sample copies; but there have been numerous occasions when we have had excellent advice from our colleagues which has saved us a great deal of trouble.

The same sort of service has also been supplied on occasion by the executive office of the Board, and I think that this might be developed to a very valuable extent, if the Board's resources were increased, particularly by making use of contacts through the National Committees of Physics in somewhat remote areas such as South America, to get reliable information of useful publications in physics.

The Board has been very useful to us in a few instances by approaching publishers with a request for copies of their periodicals for abstracting, and the extra weight given to our request by the official backing of the Board has been unmistakable. The best that the Board could achieve in one case was a 10% reduction in subscription rate, but this represented no mean concession from this particular publisher!

The discovery, appraisal and acquisition of regularly appearing periodicals, despite their ever increasing number, presents fewer problems compared to the perplexities raised by the great variety and number of so-called "non-periodical" publications and "semi-published" documents. Even if we set aside such publications as books and monographs by single authors, and dissertations presented for university degrees, whose characteristics are fairly well established, we are still faced with a formidable volume of new publication each year produced in ways which may differ widely from the customs associated with orthodox periodical publication. The most important examples are the proceedings of scientific conferences, large and small, whose contents vary widely in importance and originality. The Board has devoted a considerable amount of discussion to these publications, and though it is no part of its function to attempt to regulate the forms of primary scientific publication, it has nevertheless been driven to make some recommendations to the International Union of Physics which it believes would be of substantial value to the scientific world generally.

To assist its member journals in abstracting these publications, the Board has begun to organize the collection and circulation of periodic lists of nonperiodical publications. This experiment has started only recently, and the proper form and a common understanding among the participants has not yet been worked out. Nevertheless, it clearly has most valuable potentialities, if it can be developed in the direction of pooling the serious professional assessments of the member journals and the Board, and not degenerate into an encyclopaedic but undigested list.

Before the formation of the Board, *Physics Abstracts* had been regularly receiving page proofs of several British physics periodicals (arrangements made with the American Institute of Physics are described later). No attempt had been made to acquire the same privilege from continental publishers. We now receive proof copies of some 40 physics periodicals from France, Germany, Holland, Italy, Austria, Switzerland, Sweden, and Russia, as a direct result of approaches made either by the Secretary of the Board, or by the member journals on behalf of the Board, and in return have made available to the other member journals proof copies of about a dozen British periodicals. This wholesale interchange of proofs could almost certainly not have been achieved without the intervention of the Board, since in many cases it needed a personal approach to the publishing organization to explain the purpose of the request and to give reassurances as to the standing of the recipients. In the case of one publishing organization, which appeared to be withholding its permission unreasonably, the Board and the member journals considered the possibility of conducting, with due publicity, a boycott on the abstracting of the periodicals concerned; whether this would have been effective cannot be said, for the controversy was founded on a misunderstanding, and it was eventually happily resolved.

In the provision of proof copies of physics periodicals, therefore, the Board and the other member journals have been of definite benefit to *Physics Abstracts*. The proofs are received anything from a few days to many weeks before the published periodical, and therefore vary greatly in usefulness; further, a few of them are so incomplete and subject to alterations and re-arrangements before publication as to be unusable. A little further organization on the part of the member journals would extend the value of this service. All the British proofs are distributed through the *Physics Abstracts* office, where they are inspected before being passed on, so that deficiencies can be remedied without delay; identification details are added, if necessary (since proofs are often not titled), and post-proof alterations of which the editors inform us are passed on to the member journals. This is rendered easier by the fact that all these British periodicals are published in London; but in any case such an inspection near the source is thought to be worth a one- or two-day delay. Our continental proofs come from a variety of sources, and in any particular instance of deficiency we are usually uncertain where to enquire. It would, I believe, be well worth while for all proofs to be "vetted" before distribution by the responsible member journals, since they need to make the inspection in any case for their own purposes.

Abstracting from proof copies of periodicals has its attendant risks, and for *Physics Abstracts* we take the precaution of confirming the essential details of the articles from the published copy of the periodical just before publishing our abstracts. Occasionally we find an article has been substituted for another, more frequently that a correction has been made in the spelling of an author's name. These occurrences are rare enough, and their very rarity makes it unrealistic to expect the editor to inform us unfailingly when he makes such a change after sending the proof. This check should need to be done only once for each proof, but at present each member journal must make its own.

Arrangements with the American Institute of Physics had taken the form, before the formation of the Board, of sending by air mail clippings from the proofs of the titles and abstracts of the articles in their periodicals, since to send the whole text by air mail would have been prohibitively expensive. These "proof clippings" have proved very satisfactory to *Physics Abstracts* in the great majority of cases, for which credit must be given to the editors of the journals and to the staff of the Institute responsible for making them up. For a period of about three years, the benefit of this service rendered by the Institute was extended to the *Bulletin Analytique* by forwarding photocopies of the clippings within about three days; these were thus received in Paris almost as soon as, and at lower cost than, a separate supply. It is believed that they were useful to the *Bulletin Analytique,* which publishes short indicative abstracts, but they were unacceptable to the *Physikalische Berichte*.

We are now receiving direct from Moscow proof copies of ten Russian physics periodicals, in return for which we send proofs of British periodicals. This exchange has arisen out of the agreement reached with Professor Serpinskii, of the Russian Institute of Scientific Information, at the meeting of the Board held in August, 1956. Unfortunately, up to the time of writing, it has been our experience that these proofs arrive, in general, after the published copies purchased through ordinary commercial channels. It is to be hoped that this is due only to initial difficulties, as needless to say proof copies are only valuable if they are received well in advance. Before this service was inaugurated, single proof copies of these periodicals were sent to the office of the Board in Paris; there they were microfilmed and sent to the member journals. I regret to say that this painstaking service went for nothing as far as *Physics Abstracts* was concerned; the impossibility of using microfilm in its original form, and the expense of making enlargements, led us to discard them in favour of the published copy. Microfilm seems to us only to be justified for abstracting purposes pure and simple when the original is unobtainable.

There has been perennial discussion at meetings of the editors on the question of adopting a uniform style for the details of abstracts—the form of journal reference, and so forth—and unifying the systems of classification and indexing used for the abstracts. Progress has been very slow, no doubt because none of the participants had a strong conviction that any very good purpose would be served. It was evident that no one of the existing systems of classification is outstandingly better than the others, and insofar as the readers of the abstracting journals on the whole consult only one publication, it would not seem to concern them whether the classification system of the others is different. Nevertheless, it was thought to be worth attempting to reach agreement on a common order of subject "chapters" for the arrangement of the abstracts in the individual issues of the journals, as a first step towards a possible agreement on more detailed subdivisions. Such an arrangement of subjects, if it were reasonably sensible, might become fairly widely adopted for a number of purposes in the world of physics, and play quite a useful administrative role. I do not think there is any very sanguine belief that agreement on and adoption of a detailed "classification" system in the documentation sense would be achieved; it was simply an attempt to see whether we could agree to arrange our subjects in the same general order. This very modest attempt is rather crucial, because if agreement cannot be reached at this level, it is clearly waste of time to discuss any more ambitious scheme. The present situation is that agreement in principle has been reached on an order of subject "chapters" somewhat similar to that of the Universal Decimal Classification, and we want to-see whether the Federation International de Documentation will consent to make some corresponding revision of the U.D.C. The changes in the subject groupings of the member journals, if they materialize, will take place only slowly, but it seems hopeful that they will gradually fall into line with each other.

We have discussed two other small details of the same character; the adoption of a common set of abbreviations for the names of periodicals; and of a uniform system of transliteration of Cyrillic characters. Here again, no short-term benefit to the abstracting journals can be expected, and the stimulus is partly altruistic—if the abstracting journals can give a lead, there is an increased possibility that uniform codes for these two matters may eventually be adopted throughout a considerable part of the scientific world. The former of these may not be very important, and there are likely always to be local variations; but the standardization of a system of transliteration of Russian authors' names, or at least the reduction of the number of alternative systems in use to no more than two, is rather an urgent matter for scientific documentation, even if only a minor one. Authors' names get into circulation in a variety of ways, and if the transliterations are not the same, the differences persist as the names are copied from reference to reference; the abstracting organizations notice the discrepancies in their indexes, but in many cases have no means of determining the original form of the name. The I.C.S.U. Abstracting Board and its member journals are agreed on the value of adopting a uniform system, and I hope before long they will have some proposals to announce.

The Publications Committee of the International Union of Physics has expressed interest in these discussions of the Board, and the way seems favourable for the adoption by the Union of conclusions reached unanimously by the Board on the above subjects, for recommendation to the National Committees and the editors of physics periodicals. This is not within the province of the Board, but it gives encouragement to our discussions.

The noticeable feature of this survey is the absence of any move by the abstracting organizations themselves to share the actual work of abstracting. It will be recalled that the Board at its inception forswore any intentions of imposing such co-operation on the member journals; and I think their failure to take the initiative, though perhaps due partly to the natural conservatism of well-established organizations, is also a significant comment on the potential value of intimate multilateral working on an international scale, given the situation of three or more independent organizations producing journals in different languages. It is disappointing that this matter has not been discussed, as an examination on a practical basis of the possible scope and the difficulties to be anticipated would be illuminating; in its absence, we have only our own opinion to go on.

The conditions are not very favourable for the spontaneous growth of work-sharing, for there is more to the business of abstracting than the simple matter of composing the abstract, and an abstracting organization must work systematically as a fairly close-knit team. When we survey the differences in the basis and standards of selection, appraisal, classification, type of abstract, and indexing between ourselves and our colleagues, it is not surprising that our general feeling is that, if the original material is available to us, it is more satisfactory to abstract it independently. It may be said that these differences are anomalous, as all the member-journals are supposedly doing the same job for similar groups of people; but, as this conference will doubtless demonstrate, there are as yet few generally agreed criteria for the best form which an abstracting service should take, and, in the present situation, some variety in principles and presentation is no bad thing. Whether this is accepted or not, the limit of direct cooperation under present circumstances would seem to be the translation of abstracts of selected journals provided by our colleagues; bearing in mind the pitfalls of translating concise scientific statements without benefit of context, and the additional requirement of indexing, this course has not commended itself to us.

I would make a possible exception, at the present time, for abstracts of Russian articles; as more and more Russian literature appears, it becomes increasingly difficult to cover it in the orthodox way, because of the relatively small number of competent physicists who can read Russian. In these circumstances, the translation of a French or German abstract may well be the only alternative to having no abstract at all; and a more systematic deployment of our combined resources would probably be worth organizing. A small step in this direction was taken in 1954, when it gradually became apparent that the only copy in the Western world of the 1953 volume of the Russian *Zhurnal Eksperimentalnoi Teoreticheskoi Fiziki* was lodged in London; *Physics Abstracts* was able to supply microfilm copies to the other member journals of the Board, and we also supplied copies of our abstracts, for translation if that proved the easier course. This experiment was of course unpremeditated, and was made more on the grounds of inaccessibility of material than of difficulty of abstracting; further experiments would need to be arranged in detail in advance.

From the foregoing account it will appear that the most fruitful activities of the I.C.S.U. Abstracting Board have been in fields connected with the investigation and acquisition of material for abstracting; if circumstances remain similar to the present, and there is no closer integration of the organizations producing the abstracting journals, I expect this to continue to be the case. It is not my concern here to consider what far-reaching changes in our traditional methods may be forced on us by the rapid development of science, or what greater measure of responsibility the Board may be called on to take up, but I would like to see it extend very substantially the activities it has already developed, and in so doing work out an articulate corporate policy towards scientific publication which will have a certain degree of regulating effect. In our discussions we have discovered much common ground, and the collective opinion of the Board deserves to be heard in the interests of the individual scientific reader, whose problems are but a microcosm of those of the abstracting organizations.

In concluding my account of the benefits derived from the I.C.S.U. Abstracting Board, I should not omit to mention that the periodic meetings of the Board have afforded most profitable opportunities for discussion of matters of common concern with my colleagues of the member journals, which would otherwise not have occurred.

International Cooperative Abstracting on Building: An Appraisal

A.B.AGARD EVANS

GENERAL

This is the story of an idea, worked over by Working Parties, hammered out in Committees, adopted by a representative conference and put into practice by quite a large number of countries. It has not achieved the results expected of it. Why not? Assuming the original idea to be basically sound, what are the causes of failure? Can they be eliminated? The paper is brought to the Conference in the hope of discussion with fresh, informed and impartial minds; that the lessons learned in this field may help pitfalls in other fields to be avoided and in this way may make a contribution to technological documentation.

THEORY

The problem in 1948 was to set up machinery whereby the results of world experience and research in building might be most efficiently and satisfactorily documented for practitioners. The normal pattern of service had hitherto been for the library (or documentation centre) attached to a building research institute to subscribe to a large number of periodicals from a number of countries; to assemble books and pamphlets; and to prepare abstracts of a selection of the material thought to be of value to the research workers and other members of the employing organization. Thus there were such publications as *Building Science Abstracts* compiled by Building Research Station, Watford, England; *Ministry of Works Library Bulletin* compiled in London; *Documentations Techniques,* of l'Institut Technique du Bâtiment, Paris, and sundry others. For each such publication there was a team reading between 200 and 600 periodicals in eight or a dozen languages (some of which were relatively unfamiliar and therefore difficult to read easily and to assess the value of the paper). The cost of such a service is high and beyond the means of organizations in smaller countries.

A.B.AGARD EVANS Chief Librarian, Ministry of Works, London.

The various problems were brought to discussion at a hard-working conference in 1949 in Geneva under the auspices of U.N. Economic Commission for Europe, where plans were formulated on a practical basis and with unanimous agreement. The details were worked over and adopted at a Conference in Paris in 1950, at which the International Council for Building Documentation was formed.

So far as the abstracts were concerned, the Working Plan was briefly as follows:

(a) Each member country should have a National Building Documentation Committee;

(b) The National Committee should be responsible for covering the literature of its own country, within a defined range of subjects in the building field;

(c) Each country should prepare abstracts of its literature and provide a translation of the abstracts in a major language, i.e., English, French or Russian;

(d) These abstracts would be published by printing on cards of uniform size, with a standard bibliographic reference and a common classification (Universal Decimal Classification governed by a special selection schedule);

(e) Three copies of the abstracts would be supplied free of charge to each National Committee; further copies would be available at reasonable rates;

(f) The literature of countries not participating in the Council was to be covered by members by agreement;

(g) All abstracts should be free of copyright.

By these means it was intended to provide in each country a collection of abstracts on cards covering the most important parts of the world literature, selected by native experts, working in their own language and familiar with the authenticity of the original material. The further dissemination of the information within the respective countries was the responsibility of the National Committees. It was thought that it would be done principally by republication in a National Abstract Journal or in the pages of one or more periodicals; or that the various documentation centres in the country should subscribe to the original sheets of abstracts from abroad.

PRACTICE

Of the fifteen countries which agreed to cooperate, nine of them (Austria, Denmark, Finland, Italy, Netherlands, Norway, Spain, Sweden and Switzerland) have produced abstracts conforming to the standards. Even in these cases, however, there is some unconfirmed suspicion that the coverage is not 100% and the English (or French) translation sometimes falls short of comprehensibility.

Belgium contributes an analysis of the national literature, but in non-standard form. Germany produces abstracts in the prescribed form, but only in German and the native German articles are not segregated from those of other countries.

In the U.K., where abstracting services have long been well established, certain modifications in classification and in the size of the abstracts (to enable them to be cut and pasted to standard cards) have been made, but there has been no serious attempt to implement the international recommendations.

In France, the abstracts are produced in the prescribed form but printed on thin paper and not on cards.

The U.S.A. has not been participant. Japan and Jugoslavia have taken part in the discussions on classification, but have not (so far as the writer is aware) produced the national quota of international abstracts. Indonesia has produced some abstracts in the international form.

On the whole, therefore, after seven years, the arrangements have failed to succeed.

REASONS FOR FAILURE

I believe the plan to be philosophically sound. In practice, it has not worked. The achievement of a unified pool of abstracts of world literature in the building field is obviously desirable. The present spectacle of several centres expensively achieving a partial result and each going over the same ground in doing so is deplorable. Because the various abstract journals do cover so much common ground, it is not easy to collate them and thereby derive additional benefit from the others' efforts.

(a) The first obvious reason for failure is financial. The abstract form adopted is expensive to print in the small numbers that are required. A high proportion of the whole has to be given away (in exchange for other abstracts, it is true) and the remainder (limited to national abstracts) has a very limited market. In the case of the smaller countries the total sum involved is relatively insignificant and can be met by members of the National Building Documentation Committee, putting their hands in their pocket, if need be. In larger countries whose contribution may be as high as 1000 abstracts, the expense is a budget item, and a difficult one to sustain. The return, consisting at present of abstracts from only the smaller countries, is of slight counterbalancing weight.

(b) The second reason, particularly in the larger countries, is vested interest. There is no question of setting up a brand-new organization for the job. It must be done by existing organizations, whose primary duty is to their own membership.

In the U.K., for example, there are three principal building documentation centres: viz., Building Research Station, which has published *Building Science Abstracts* since 1928; Royal Institute of British Architects which has brought out a *Library Bulletin* for over 20 years; and Ministry of Works, which has produced a *Library Bulletin* since 1944.

All these publications have developed a form designed to serve the needs of their own readers. The provenance is world-wide and the subject coverage is that required by the organizations. The essence of analytical journals is their continuity; they cannot be drastically altered in order to serve a problematic scheme.

Therefore, until an international scheme is proven to be a satisfactory substitute, they must continue in their present form. Each represents a major effort by the staff employed on them and there is little surplus labour available for additional tasks. It would be possible for some appointed person to select abstracts from these publications and to republish in international card form. But the cost would not be less than £2,000 a year, with no visible return. This proposal was canvassed, but adequate support was not available. (It is practicable to supply the abstracts, if a publisher could be found: see below.)

Meanwhile, *Building Science Abstracts* adopted the practice of adding UDC Classification; the *M.O.W. Library Bulletin* was produced in double-column form so that the abstracts could be pasted onto cards and interfiled with the orthodox international cards.

Similar considerations affected the French. In Germany, the abstracts were required to be financially self-supporting: a great and praiseworthy effort has been made by producing them in international form; but the customers required the analysis to include foreign literature and arrangements could not be stretched to provide translation and publication in English or French.

No effective body seems to have appeared in the U.S.A. to participate and, in view of the unsatisfactory position to date, there would seem to be little incentive to do so in the present circumstances. The abstracts produced by the Building Research Institute could be developed on the international lines, given the right conditions.

(c) A third factor of importance is the need for central editing. A guiding hand, rather than a dictatorship, is required if a satisfactorily uniform standard of paper is to be selected for abstracting. Additionally, some of the translations would benefit from verbal editing.

(d) The fourth factor is marketing or availability. In order that an institution may obtain all the various abstracts, it is necessary first to discover who publishes what at how much and then to enter quite a large number of subscriptions in various currencies. They are not generally available through international

booksellers. The frequency of publication tends to be erratic, thus making control more difficult.

CONDITIONS FOR SUCCESS

Considerable progress has been made in a plan for Bouwcentrum, Rotterdam, to publish a consolidated edition in English. This would relieve individual countries from the need to publish internationally; provide the editorial supervision; solve the marketing problem.

Financing the scheme presents the difficulty that the national contributions required are comparatively trivial. Many of the National Building Documentation Committees are consultative, rather than executive and, since the work is done by a few member organizations, the creation of working funds has not been seriously tackled, because there has been no recognizable need for them. A committee is a hopeless organization for Executive action. It is obviously desirable for each country to throw up a Building Documentation Centre, with responsibility for International liaison and national action contributing thereto; in practice, it is not so easy to find a volunteer to shoulder the burden alone, nor to achieve a practical and equitable contribution of money and effort from the other members.

However, a solution of the problem is required, if the worthwhile objective is to be achieved. The matter is of additional importance, since building documentation is developing actively in Asia and the Far East. The cooperation of USSR and the east European countries is a distinct possibility. Hungarian Technical abstracts and the Polish Institute of Housing already produce some material of interest and it would not be difficult for them to produce national abstracts in the building range from among the technical abstracts already being prepared.

Cooperation and Coordination in Abstracting and Documentation

OTTO FRANK

ABSTRACT. The following report is divided into three main parts: (1) a description of examples of coordinated documentation in Europe and the experiences derived therefrom; (2) guide lines for coordinating methods derived from these experiences; and (3) cost savings by coordinated documentation.

1. DESCRIPTION OF EXAMPLES OF COORDINATED DOCUMENTATION IN EUROPE AND THE EXPERIENCES DERIVED THEREFROM

1.1. ELECTRICAL ENGINEERING

In Germany documentation bureaus of large companies and administrations in the field of electrotechnics, e.g., the AEG [Allgemeine Elektrizitaets-Gesellschaft—General Electric Company] in Berlin and Frankfort on the Main, the Siemens-Schuckert plants in Erlangen, Siemens and Halske AG in Munich, Standard Elektrik in Stuttgart, Alldephi in Hamburg, and the Central Telecommunications Bureau in Darmstadt, have gone through the technical literature and have reproduced the titles of publications in the form of bibliographic cards or lists and distributed them to the associates of their companies or administrations.

The large amount of time and money that has gone into this project has prompted deliberations on how, through a joint venture, documentation could be made better though less expensive.

The simplest way, i.e., the direct exchange of bibliographic cards and lists among companies, did not prove practical because many companies were afraid that this might enable other companies to gain insight into their research and planning. All participating companies, however, showed a readiness to

OTTO FRANK Allgemeine Elektrizitäts-Gesellschaft Patent Buro, Frankfort on the Main, Germany.

entrust the distribution of their bibliographic evaluation to a neutral organization.

The Association of German Electrical Engineers (VDE) in Frankfort on the Main declared itself willing to take on this task.

In the negotiations, however, it became evident that a lowering of the cost could not be achieved by an exchange alone, because it would not decrease the work of the individual companies. The solution was found in a division of work agreed upon by the participating concerns. The roughly 300 periodicals were divided among the cooperating companies in such a manner that no one company has to evaluate all periodicals any longer, but only a part of them. The titles of all publications considered important in this connection are sent to the Association of German Electrical Engineers, which organization sorts them and publishes them once a week in the form of “VDE Flash Reports.” All participating companies expressed themselves in favor of fast reporting because abstracts frequently do not appear in the regular periodicals until several months after the publication of the original work. They agreed that the VDE Flash Reports would report only the titles of publications without any indication of contents and without abstracts.

The price for the VDE Flash was kept low (yearly subscription—80.00 DM[1] for about 10,000 titles, i.e., less than 0.01 DM per title, in order to make it possible for participating companies to order large quantities of copies for distribution to their associates).

The savings stem mainly from the following points:

1. Every participating company needs to go through and evaluate only a small part of the technical journals.
2. In this manner, individual companies are saved the expense of writing up and reproducing their own bibliographic lists. All participating organizations ceased publication of their own bibliographic reports, evaluations being continued only in especially important branches. In the latter case, detailed descriptions of contents and abstracts are supplied.

Though only a few large companies cooperate in compiling the data, those companies agreed that VDE Flash Reports should be sent not only to the participating companies but that they should also be made available to anybody in Germany or abroad.

1.2. CONTROL TECHNOLOGY

The division Control Technology of the Association of German Engineers (VDI) and the Association of German Electrical Engineers (VDE) felt the

[1] 4.3 DM=1 dollar.

necessity to keep their members and all experts interested in control technology informed about new publications in the field of control technology. This is a difficult task for any professional man, because publications concerning control technology appear in almost all technical publications. This would require him to scan a large number of publications.

A solution was found, similar to that of the VDE Flash Reports. The division acquired a number of collaborators who declared themselves willing to scan certain professional publications and to report to the division any articles of importance to control technology. The division then compiles the titles and publishes them in a monthly periodical entitled *Regelungstechnik* [Control Technology] in the form of a periodical review. It also covers publications on automation, especially in production engineering.

In this way, subscribers to the periodical *Regelungstechnik* are kept informed about new publications and can save themselves the scanning of numerous professional publications. The periodical review can also be subscribed to printed on one side only, which facilitates inclusion in a bibliographic file.

1.3. AVIATION

The documentation organizations in the aviation field in Germany, especially those of the German Experimental Institute for Aviation and the Scientific Society for Aviation, collaborate in the Central Organization for Aviation Documentation. Every one of these documentation organizations works on one or several branches, in such a way that the technical publications are scanned and evaluated by scientists working in scientific institutes. Titles and abstracts are sent to the central organization, which compiles, reproduces, and distributes them. In this way, every individual documentation organization needs to take care of only a small part of the evaluation, but it receives from the central organization all bibliographic cards from all documentation organizations.

There is cooperation and exchange with documentation organizations of the aviation industries in France and Great Britain, e.g., with the Ministry of Supply and the Institute of Aeronautical Science in London. In addition to technical publications, the research, testing, and industrial reports published in the various countries are of special importance to aviation documentation. These can be made available to interested persons by the central organization of aviation documentation (frequently in the form of microfilms).

1.4. STRUCTURAL ENGINEERING

The Conseil International du Bâtiment has representatives from numerous European countries that have decided on a joint documentation program.

This is now being carried out. Every national member compiles the titles of works that appeared in his country in the field of structural engineering and exchanges them with all other countries. Standard form A7 (74×105 mm) has been selected for the international exchange. The titles are supplemented by a table of contents in the language of publication and (on the reverse side of the card) a table of contents in English or French. The countries have granted each other the right to use the titles and the tables of contents for reproduction in original form or for complete or partial translations thereof.

In Germany the abstracts received in exchange by the Documentation Organization for Structural Engineering of the German Construction Center in Stuttgart as well as the abstracts based on work done in Germany proper, especially those selected from German publications, are processed and published in the *Schrifttumskartei Bauwesen* [Bibliographic Index—Structural Engineering].

The scanning of the professional journals is assigned to a number of associates who send the titles selected by them to the Documentation Organization for Structural Engineering.

Many of the documentation organizations now operating in Germany work in areas related to the problems of structural engineering, e.g., the Documentation Organization of the Deutsche Bundesbahn [Railroads of the Federal Republic of Germany] in Offenbach on the Main, the Research Society for the Highway System, the German Commission for Steel Construction, and the German Association for Housing and Urban-Construction and Clearance Planning in Cologne. The Documentation Organization for Structural Engineering has concluded agreements with these organizations concerning the exchange and free use of titles and tables of contents. This permits effective evaluation of works published.

1.5. ECONOMIC SCIENCES

The library of the Economische Voorlichtingsdienst van het Ministerie van Economische Zaken (Economic Information Service of the Ministry of Economic Affairs) in The Hague processes and coordinates the documentation of publications (books, brochures, essays in periodicals and newspapers) in the area of political economy and industrial engineering. The titles and abstracts are selected jointly by:

de Bibliotheek van de Economische Voorlichtingsdienst.

de Bibliotheek van het Ministerie van Sociale Zaken (Library of the Ministry of Social Affairs) in the Hague.

de Bibliotheek der Nederlandsche Economische Hoogeschool (Library of the Dutch Institute of Economics, in Rotterdam.

In every organization experts check and process the publications assigned to them (altogether 2,500 periodicals). The abstracts contain not only bibliographic data but also tables of contents (100–200 words) in the original language (English, French, German, Dutch); abstracts from other languages (Slavic, Scandinavian, Romance languages) are given in Dutch. In the library of the Economic Information Service in The Hague the abstracts are written on transparent paper and reproduced by the diazo process. About 20,000 abstracts are being produced yearly, which are distributed to subscribers. A selection of these abstracts is published in *Economic Abstracts,* a bimonthly collection of abstracts.

In this manner, the users of this documentation service are saved the time normally spent in checking numerous periodicals and making out bibliographic cards.

1.6. INSTITUTE FOR DOCUMENTATION IN THE GERMAN ACADEMY OF SCIENCES IN BERLIN

The German Academy of Sciences has a very complete documentation program in the numerous fields of the natural sciences, technology, and the social sciences, which has developed into a coordinated and cooperative documentation network.

The documentation organizations of the individual branches, e.g., production engineering, measuring technique, control and regulation technology, foundry practice, construction materials, synthetics, optics, geophysics, fibrous materials and textile technology, and labor economy, have formed an association whose most important characteristic is a unified methodology (form, content, reproduction and distribution of the bibliographic cards) and a close delimitation of the various fields of work to prevent duplication of effort and overlapping).

The standard professional publications of all countries, especially also of the eastern European countries, are scanned in the individual documentation organizations, [whereupon] the selected titles and abstracts are processed according to standard directives.

One part of the documentation organizations takes care of the reproduction and distribution to subscribers, another part supplies the manuscript sheets to the Academy Publishing House in Berlin, which takes charge of reproduction and distribution.

The Institute for Documentation of the German Academy of Sciences in Berlin is responsible for overall organization.

Altogether some 100 different information services publish cards or lists for the various fields, totaling over 100,000 title cards per year. The total number of cards sold is about 15,000,000. These cards are received by about 5,600 institutions and individuals, representing about 6,000 individual subscriptions.

(It may be noted that among these bibliographic services there is one which deals specifically with questions of documentation and regularly announces the publications in this field.)

Besides that, the institute publishes several collections of abstracts: the *Central Chemical Journal,* the *Central Technological Journal,* and the *Central Agricultural Paper*.

2. DIRECTIVES CONCERNING METHODS OF COORDINATION

Collaboration and coordination in the field of documentation is complicated by two facts:

1. In many instances there is opposition to cooperation, e.g., opposition based on fear that privately worked out documentation practices and forms will have to be abandoned in favor of a new, coordinated documentation. This is a psychological problem.
2. Those organizations that are willing to combine their documentation work with that of other organizations working in the same professional field, are not sufficiently familiar with the methods which would enable them to do this more simply and economically. This is an organizational problem.

To solve both of these problems the German Society for Documentation set up a Commission on Methods of Coordination, which is under the direction of the author of this report. The commission's task is to work out directives according to which the various professional circles, e.g., technological-scientific associations, research institutes, companies, and technical administrations, can arrange their documentation in a standard and economic fashion (see examples in Sec. 1). These directives will deal with the following individual questions, which are of importance to coordinated documentation.

2.1. ANALYSIS OF THE STATE OF DOCUMENTATION IN THE PROFESSIONAL FIELD

First of all, it is important that an exact picture be obtained of the existing documentation. This involves the following specific problems:

2.11. Determination of the total number of publications appearing each year

This determination should permit an assessment of the work load to be expected, e.g., with respect to the number of books, periodicals, articles, reports, patent publications, etc.

2.12. Determination of the bibliographic publications

This determination should provide insight into already existing bibliographic works that can be used for coordinated documentation, e.g., title indexes, bibliographies, collections of abstracts, indexes of abstracts, and documentation sections of journals.

2.13. Determination of the existing documentation organizations

A compilation should list not only public documentation organizations but also documentation organizations operated by commerce and management, e.g., industrial concerns at home and abroad.

2.14. Determination of the organizations considered for cooperation

In this category belong scientific organizations, scientific and research institutes, professional libraries, companies, and individual experts.

2.2. ORGANIZATIONAL QUESTIONS

The following questions are of importance in connection with the organization of a cooperative program:

2.21. Arousing interest in coordinated documentation

This preparatory work by personal contact or by written questionnaires should indicate whether there exists sufficient interest to warrant preparations for discussions.

2.22. Calling a conference

In order to avoid identification with commercial interests, the invitation should be extended by a neutral party, if possible. It should be accompanied by highly detailed information on points to be discussed; e.g., an analysis of the type described under 2.1, and a plan for coordination.

2.23. Discussion, resolution, agreements

The discussion should clarify the possibilities outlined under points 2.3 to 2.5 and lead to resolutions regarding the design of a coordinated documentation. This includes the appointment of persons and institutions to take charge of various tasks, either in directive or executive capacities. Finally, it also includes agreements on the form of the coordinated documentation: cards or lists, format of cards, order of title information, tables of contents, abstracts, classification characteristics, and settlement of language and translation problems.

2.3. DIRECT EXCHANGE OF INFORMATION DATA

Direct exchange between the participating organizations is the simplest mode of cooperation. One can differentiate between:

2.31. Uncontrolled exchange

In this case all participating organizations exchange with one another, every organization providing as many copies, e.g., of bibliographic cards or lists, as are needed by the exchange partner.

The advantage of this method is that in this way the individual organizations receive information which they did not find themselves. The disadvantage is that this simple exchange leads to numerous duplications, as a result of which the individual exchange partners are not saved much work.

2.32. Controlled exchange

In this case the bibliographic evaluation work is distributed among the participating organizations according to a certain plan. This can be done, for example, according to professional fields, every partner taking on the evaluation of a certain division. The division according to periodicals is simpler, every partner scanning and evaluating a certain number of publications. Furthermore, in the international field a division according to geographic criteria is possible. In this case each country takes charge of the evaluation of publications appearing within its borders.

The advantage of this method lies in the fact that the amount of work for the individual members decreases in a manner inversely proportional to the number of exchange partners. Assuming six exchange partners, for example, every individual needs to do only one-sixth of the evaluation work, but in exchange he receives the results of all six partners, i.e., the entire documentation.

2.4. COMPILATION IN A CENTRAL ORGANIZATION

In this case the partners do not exchange evaluation results directly with each other but transmit them to a central organization. The following questions need to be cleared up in this connection:

2.41. Division of the professional fields

Professional fields must be delimited as closely as possible. Furthermore, one must determine which allied fields are to be included and to what extent the publications in the allied fields should be evaluated. In addition, cooperation with documentation organizations in these allied fields may be worth con sidering (compare Sec. 2.6).

2.42. Format questions

Format, card content, entries of characteristics of order, etc., must be uniformly established.

2.43. Division of work

The partners must agree on the division of work; for example, on the publications which every partner accepts for evaluation.

2.44. Kinds of publications

Agreements must be reached on what types of publications should be utilized; for example, only periodical articles or also books, patent publications, directives, reports, etc. It must be agreed upon whether all publications should be evaluated or whether a selection should be made, for example according to size, scientific value, language, common or solely specific importance.

2.45. Setting up a central bibliographic index

On the basis of these findings one may decide to set up a bibliographic index at the central organization, which will be available to all cooperating organizations.

2.5. DISTRIBUTION OF THE INFORMATION

In many cases coordinated documentation will lead to the distribution of information to partners in the form of bibliographic cards, and indexes, etc. The following points need to be clarified in this connection.

2.51. Form of information

The most effective form is that of bibliographic cards, which recipients can incorporate in their own bibliographic indexes. If the number of participants is large, lists or booklets may be more practical because they are cheaper. The most suitable reproduction process must be selected.

2.52. Type of Release

In this connection, one must decide upon the speed of reporting and also on the periodicity of the publication.

2.53. Selection

It is conceivable that although all information collected by the central organization is incorporated in a central index, only a part of it is ever distributed.

In such instances the recipients receive a selection of all information material but may request additional information from the central organization in specific instances.

2.54. Division by groups

If not all partners are interested in all individual fields of the coordinated documentation, then it is practical to set up a division by subject groups so that every partner can receive the group of interest to him—if need be, in a greater number of copies—for distribution in his enterprise.

2.55. Circle of recipients

It must be agreed upon whether the information should be distributed only to the partners participating in the compilation or whether it should be available to all experts. In this case one may do well to consider publication by reproduction or printing, and an agreement should be reached as to whether the publication is to be handled solely by the central organization or whether cooperation with a publishing house should be sought. One should determine whether the publisher of a technical journal might be willing to publish the information as part of his journal, e.g., as a review of periodical literature.

2.6. COOPERATION WITH DOCUMENTATION ORGANIZATIONS IN RELATED FIELDS

Even the most careful division of professional fields will not completely prevent overlap with the documentation work being done in other fields. This is especially true of fields that cut across other fields. A good example of this would be control technology. Control technology can be considered a field in itself and documented as such. However, publications on control technology are always important to many fields and are evaluated by the documentation organizations of these fields, e.g., power supply, machine building, electrical engineering, and production engineering (in automation). Similar things are true of the basic sciences, e.g., physics and chemistry, which can be documentated as separate subjects, but they have numerous ties with the applied sciences, e.g., with aircraft construction and synthetics industry.

It is advantageous to agree on an exchange of information data with documentation organizations in related fields. This may be limited to information of interest from a technical viewpoint. For example, a documentation organization in electrical engineering may transmit all available information on structural engineering (e.g., on house installations) to a documentation organization in that field. The latter in turn makes available whatever information it has on electrical engineering. Agreements must also take in consideration copyright questions; for example, the free distribution of the information received in the exchange and the right to incorporate it in the reproduced or printed information services.

3. SAVINGS THROUGH COOPERATIVE WORK

It is difficult to calculate exactly the savings of documentation methods because the controlling factors are very different in nature. In any comparison of the costs of ordinary documentation with those of coordinated documentation, the absolute values are therefore less important than the relative values.

The cost comparison made below is not based on the point of view of large documentation organizations that have at their disposal sufficient scientific personnel, highly perfected technical equipment and accessories, as well as ample financial means. Medium-sized and small documentation organizations and those in industrial concerns often lack these means. To them economical operation is essential, while cooperative work is of special importance. The same holds true especially for medium-sized and small companies, which frequently cannot even afford assigning bibliographic evaluators on a full-time basis.

3.1. COST OF BIBLIOGRAPHIC CARDS IN LOCAL EVALUATION

The following costs, which were obtained through thorough and reliable studies and comparisons, may be considered average for Germany:

Preparation of a bibliographic card with title and bibliographic sources of a publication	1.00 DM
Preparation of the same bibliographic card with added subject index	3.00 DM
Preparation of a bibliographic card containing an abstract or whatever facts are needed for mechanical selection	10.00 DM or more
Cost of classifying a title in decimal classification or in another system	1.00 DM

In the above the main part of the cost lies in the high caliber of mental work required in the selection of the titles, the preparation of the table of contents or abstract, and the classification. The costs for writing and material total about 10% of the total and therefore are not considered in the following analysis.

3.2. SAVINGS DUE TO DIVISION OF WORK

Let us assume that 1,000 publications which appeared in a given period in 100 technical journals in a certain field must be included in the bibliographic index. If a subject index is added to the title information, then the cost for the writing of the 1,000 title cards will amount to 3,000.00 DM.

If two partners agree to cooperate in the documentation so that either partner scans 50 publications and produces 500 bibliographic cards, then the cost of either partner is reduced 50%, i.e., to 1,500.00 DM. Nonetheless, the addition to his files is the same as in the first case, namely, 1,000 bibliographic cards. That is, not only the 500 he produced but also the 500 cards prepared by his partner.

If five partners agree on joint documentation so that each of the five partners scans 20 publications and supplies 200 title cards, then the costs for every partner are reduced to one-fifth, i.e., to 600.00 DM. Again every partner receives all the cards in the exchange, as in the first case.

(This does not take into consideration the costs of reproducing the cards; for example, by the blueprinting process.)

The larger the number of exchange partners, the lower the costs for the individual partners. From this we can derive the following rule: The documentation is the more economical the larger the lots in which the individual bibliographic cards are being produced. This is a known fact. But it is also a fact that little use is being made of it and that many documentation organizations that evaluate publications produce cards only in single copies (or, at any rate, only in small lots), which are written out and included in the bibliographic file.

3.3. COST OF BIBLIOGRAPHIC CARDS FROM REVIEWS AND REFERENCE INDEXES

The above-mentioned rule of the economy of documentation is confirmed most clearly when the bibliographic cards are reproduced, i.e., printed, in very large numbers. One can compute very easily that the title information in the *Engineering Index,* the *Chemical Abstracts,* the *Central Chemical Journal,* the *Science Abstracts,* etc., cost an average of 0.05 DM. If these bibliographic items are cut out and pasted on to cards, then the price for the finished bibliographic card amounts to 0.10 DM, most of these cards containing an abstract or a table of contents. This means that a bibliographic card made up in this fashion requires only 1/30 to 1/100 of the cost of self-compiled bibliographic cards.

An examination of reviews appearing in card form, e.g., the card form of the *Engineering Index* and of the *Bibliographic Services* appearing in card form in Germany, shows that these cards cost an average of 0.10 DM, which corresponds to the cost of a bibliography card prepared by the cutting out and pasting up of abstracts.

3.4. IMPORTANCE OF STANDARD CLASSIFICATION SYSTEMS

Every cooperative venture in the field of documentation must have its origin in uniform foundations, e.g., with respect to the size of the bibliographic cards and the type of title arrangement.

Especially important in this connection is the classification system. If several exchange partners employ different systems, then every exchange partner on receiving the cards must classify them according to his own system. If these exchange partners use the same classification system, then this work is not necessary, and the cards can be inserted in the bibliographic index with little trouble.

This explains why all examples of coordinated documentations listed in Sec. 1 are based on the international decimal classification. Since a very well-prepared edition of this classification (complete with comprehensive alphabetical index) is available in Germany we need make no further mention of the classification system in the examples listed. In no instance was there any need to set up a special classification system. The simple agreement that the publications to be evaluated were to be classified according to the decimal classification was sufficient. This had the added advantage that in Germany (and in other European countries) most of the important technical journals classify the original publications by using the numbers of the international decimal classification. These numbers can be copied on to the cards. Thus, when the cards are to be written up, one no longer has to classify them.

It is therefore obvious that the recipients of the bibliographic services described in Sec. 1 also use the decimal classification to systematize their bibliographic indexes. The cards therefore need not be classified, but can be filed in the bibliographic index by clerical personnel.

3.5. POSSIBILITIES OF COORDINATED DOCUMENTATION USING HAND- AND MACINE-PUNCHED CARDS

Hand-punched cards (punched or notched) are being used in increasing measure for documentation purposes, while the possibilities of using machine-punched cards remain restricted to few large documentation organizations because of the cost of the machine equipment. It probably will remain so restricted in the future.

Experience gained in the use of hand-punched cards shows that the majority of classification systems now in use were devised by the individual documentation organization themselves. Even though the international decimal classification has also been used for that purpose, experience seems to indicate that it is not always the most suitable classification system for mechanical selection because of the very long figure symbols.

Up to now hand-punched cards and machine-punched cards have not been used in coordinated documentation. As a rule, the cards used in this type of work are made out only once and used only in one single documentation office. Besides, the advantages of coordinated documentation with mechanical selection are even greater than those of standard documentation with the usual bibliographic indexes, because the preparation of a bibliographic card for a file of hand-punched or machine-punched cards is considerably more expensive (because the frequently large number of factual data must be decoded) than that of ordinary bibliographic cards. Coordination will permit the same measure of cost reduction for mechanical sorting as can be attained for ordinary bibliographic cards. As a matter of fact, it appears as though mechanical sorting can be used to full advantage only if the principles of division of labor and of coordination be applied in setting up the files concerned.

BIBLIOGRAPHY

OTTO FRANK. Problems of coordinating documentation—Documentation of electrical engineering, *Nachrichten für Dokumentation* [Documentation News], *5* [4], 187–190 (1954).

OTTO FRANK. The new periodical review in control technology, *Regelungstechnik* [Control Technology], *3* [1], 1 (1955).

HANS-JOACHIM RAUTENBERG. Problems and status of the work of aviation documentation in the Federal Republic of Germany, *Nachrichten für Dokumentation, 7* [4], 192–196 (1956).

CLARA E.MUELLER. The work of the Structural Engineering Information Office of the Research Society for Construction and Housing, *Nachrichten für Dokumentation, 2* [2], 69–72 (1951).

L.M.GIERTZ. Construction documentation in Sweden, *Nachrichten für Dokumentation, 3* [2], 76–80 (1952).

OTTO FRANK. Roles and activities of the Conseil International du Bâtiment, *Revue de la Documentation* [Documentation Review], *21* [1], 19–20 (1954).

GERDA M.VAN ANDEL. Economic documentation in the Netherlands, Rapp. 18e FID, *1* [6] (1951), Rome.

GERDA M.VAN ANDEL. Economic documentation in the Netherlands, *Nachrichten für Dokumentation, 3* [2], 71–74 (1952).

J.KOBLITZ. The organization of pointing and distributing the documentation services of the German Democratic Republic, *Revue de la Documentation, 24* [1], 31–35 (1957).

HELEN L.BROWNSON. Coordination of information services, *Library Trends, 2* [1], 127–145 (1952).

On the Functioning of the All-Union Institute for Scientific and Technical Information of the USSR Academy of Sciences

A.I.MIKHAILOV

Rapid progress of science and technology is only possible today with a well-organized information service. The advancement of scientific and technical thought has resulted in the division of knowledge into an infinite number of fields. It is virtually impossible for an individual specialist or even a group of scientists to peruse the output of publications pertaining to their particular sphere.

As the well-known British scientist Prof. John Bernal pointed out: "The situation has indeed been reached in many fields where it is easier to find out a new fact or build up a new theory than to ascertain whether these have been discovered or deduced before."[1]

At the present-day level of scientific development one and the same problem is often studied by a number of institutions in several countries simultaneously. Hence many problems of science and technology may be solved in less time with the help of an all-encompassing systematic information service.

This need has caused the widespread appearance of numerous informational publications, such as abstracts journals giving a compact concentration of the problems dealt with in specialized literature. Abstracts journals have become a vital tool in scientific research, created by progress of world culture. The need for abstracting scientific literature was recognized in Russia as far back as the middle of the 18th century and expressed by one of her eminent scientists Mikhail Lomonosov.

According to Lomonosov, scientific abstracting is hard and complicated

A.I.MIKHAILOV Director, All-Union Institute for Scientific and Technical Information, Academy of Sciences, USSR.

[1] John Bernal "Science in History," Additions to the Russian Edition, Moscow, 1956, page 681.

work the purpose of which is not merely to convey known facts and general truths, but to display *know-how* in grasping new and *essential* facts appearing in works sometimes conceived by geniuses.[2]

In the latter half of the 18th century a number of Russian periodicals regularly published abstracts of scientific works. However, special information periodicals—in the sense this term is used today—appeared in Russia only at the opening of the 19th century. In pre-Revolutionary Russia from 1800 to 1917 there appeared at different times some 50 publications which were, in whole or in part, devoted to information.

Considerable attention was paid to scientific and technical information already in the very first years following the Revolution. The publication of an abstracts journal —"Reports on Scientific and Technical Works in the Republic"—was started as early as January 1920, i.e. during the Civil War. For those days it had a comparatively large circulation of 2,000 copies.

These "Reports" printed brief summaries of scientific and technical projects — completed, still underway and those just launched. The "Reports" had the following sections: Physics, Chemistry and Chemical Technology, Geology, Agriculture, Bacteriology, Applied Mechanics, Mechanical Technology, Metallurgy and Engineering.

Later a number of publications appeared in the USSR carrying information on Soviet as well as foreign scientific and technical literature. These were prepared by publishing houses, libraries and branch information centres. As in most countries, there was still no centralized information service in the USSR.

However, only a centralized system for issuing abstracts journals is capable of ensuring a more or less total (rather than relative) coverage of information, for only such a system affords the means for collecting, systematizing and generalizing all the facts dispersed throughout the multitude of sources. On the other hand such a set up offers the advantage of avoiding duplication thereby saving on effort and means.

In 1952 the Presidium of the USSR Academy of Sciences—proceeding from national and international experience in the field of scientific information— organized within the Academy framework a special Institute for Scientific Information. Later it was re-organized into the All-Union Institute for Scientific and Technical Information (VINITI).

The All-Union Institute for Scientific and Technical Information is a specialized scientific research establishment whose object is to provide scientists and

[2] Collected Materials on the History of the Imperial Academy of Sciences in the 18th century. Published by A.Kunik. Part 11, St. Petersburg, 1865. "On the Role of the Journalist in Abstracting Works for the Maintenance of the Freedom of Reasoning," pp. 517–518, Italics by Lomonosov.

technicians with exhaustive information on all the achievements in science and technology throughout the world.

Starting with October 1953 the Institute launched the publication of an Abstracts Journal—exhaustive information reference hand-book. As distinguished from most countries where independent abstracts journals mark the prevailing trend, the USSR issues a systematic Abstracts Journal in several series, which comprises a universal encyclopedic reference manual covering all the latest achievements in science and technology.

The Abstracts Journal series constitute only the primary stage in the processing of the entire mass of scientific information. The classified facts already entered in the Abstracts Journal can further be grouped under numerous headings. This procedure sometimes helps to discover some still unformed trend in scientific development.

Various indexes—authors', subject, systematic, formulary, etc.—enter the Abstracts Journal series as component parts considerably extending the uses of these publications. The Abstracts Journal series also provide a basis for the preparation of all sorts of reference hand-books on problems of vital interest— both broad and narrow in scope.

In 1953 the Institute started publishing the Abstracts Journal in four series: “Chemistry,” “Mechanics,” “Mathematics,” and “Astronomy and Geodesy.”

In 1954 three additional series appeared: “Physics,” “Biology,” “Geology and Geography.”

In 1955 the Abstracts Journal was issued in eight series: “Mechanics,” “Mathematics,” “Physics,” “Biology,” “Astronomy and Geodesy,” “Chemistry,” “Biological Chemistry,” and “Geology and Geography.”

In 1956 the Abstracts Journal started to come out in 12 series: “Astronomy and Geodesy,” “Biology,” “Geography,” “Geology,” “Mathematics,” “Engineering,” “Metallurgy,” “Mechanics,” “Physics,” “Chemistry,” “Biological Chemistry” and “Electrotechnics.”

In 1957 the Abstracts Journal was published in 13 series encompassing natural, exact and technological sciences: “Astronomy and Geodesy,” “Biology,” “Biological Chemistry,” “Geography,” “Geology,” “Geophysics,” “Mathematics,” “Engineering,” “Metallurgy,” “Mechanics,” “Physics,” “Chemistry” and “Electrotechnics.”

In 1953–1954 the Abstracts Journal series had 107,890 entries, 209,967—in 1955; 391,481—in 1956 and 455,000 in 1957.

The overall volume of the complete edition of the Abstracts Journal in 13 series comprised over 12,000 author's quires[3] which is equal in size to 100 volumes of the Great Soviet Encyclopedia.

[3] An author's quire is 40,000 typographical characters.

As of January 1, 1958, 1,064,338 entries appeared in the Abstracts Journal series.

In 1958 the Abstracts Journal continued to come out in 13 series:

“Astronomy and Geodesy,” monthly edition. Each issue normally carries 700 abstracts, annotations and bibliographical descriptions of magazine articles, dissertations, books, patents, etc. The annual subscription price is 115 roubles 20 kopeks.

“Biological Chemistry,” fortnightly edition. Each issue normally carries about 1,300 abstracts, annotations and bibliographical descriptions. The annual subscription price is 216 roubles.

“Biology,” fortnightly edition. Each issue normally carries about 4,500 abstracts, annotations and bibliographical descriptions. The annual subscription price is 691 roubles 20 kopeks.

“Geography,” monthly edition. Each issue normally carries about 2,500 abstracts, annotations and bibliographical descriptions. The annual subscription price is 288 roubles.

“Geology,” monthly edition. Each issue normally carries about 1,500 abstracts, annotations and bibliographical descriptions. The annual subscription price is 288 roubles.

“Geophysics,” monthly edition. Each issue normally carries about 850 abstracts, annotations and bibliographical descriptions. The annual subscription price is 115 roubles 20 kopeks.

“Mathematics,” monthly edition. Each issue normally carries 800–900 abstracts, annotations and bibliographical descriptions. The annual subscription price is 172 roubles 80 kopeks.

“Engineering,” fortnightly edition. Each issue normally carries 3,500–4,000 abstracts, annotations and bibliographical descriptions. The annual subscription price is 734 roubles 40 kopeks.

“Metallurgy,” monthly edition. Each issue normally carries over 2,000 abstracts, annotations and bibliographical descriptions. The annual subscription price is 504 roubles.

“Mechanics,” monthly edition. Each issue normally carries 1,300–1,400 abstracts, annotations and bibliographical descriptions. The annual subscription price is 172 roubles 80 kopeks.

“Physics,” monthly edition. Each issue normally carries about 2,500 abstracts, annotations and bibliographical descriptions. The annual subscription price is 360 roubles.

“Chemistry,” fortnightly edition. Each issue normally carries about 3,500 abstracts, annotations and bibliographical descriptions. The annual subscription price is 756 roubles.

"Electrotechnics," fortnightly edition. Each issue normally carries 2,100– 2,300 abstracts, annotations and bibliographical descriptions. The annual subscription price is 480 roubles.

Starting with 1958 a number of the Abstracts Journal series—"Biology," "Geography," "Engineering," "Metallurgy" and "Chemistry"—apart from the complete editions are also being issued in separate reprints in booklet form covering individual sections of the former. These booklets are intended for those readers who do not need the complete editions. For instance, the "Engineering" series comes out in five booklets: "General Problems of Engineering and Machine Designing," "Technology of Machine Building, Metrology, Measuring and Control Apparatus," "Technology and Equipment for Casting," "Transport, Traction and Hoisting Equipment and Engines," and finally, "Branch Machine Building."

The "Metallurgy" series, apart from the complete edition comes out in two booklets: "Metallurgy of Iron and Steel" and "Welding."

Besides publishing the Abstracts Journal the Institute prepares selective information on various subjects. The Institute publishes "Express Information" bulletins containing information on the more vital scientific and technological problems. The bulletins are issued in a number of series, such as "Automatic Control of Production Processes," "Automobile Construction and Motor Transport," "Computers," "Public Health and Medicine," and "Forging."

The periodicity of the "Express Information" bulletins is 48 issues per year.

On the basis of the Abstracts Journal series the Institute has launched a series of monographic reviews under the general heading "Advances in Science." These reviews cover the major problems of Physics, Mathematics, Chemistry, Geology, Geography, Biology and Technology. Their intention is to systematize and generalize the achievements made in the scientific branches over a definite period of time. The reviews are issued for the benefit of scientists and specialists working in industry and agriculture. In 1957 the first two issues appeared: "Problems of Theory of Non-Linear Systems of Automatic Adjustment and Control" and "Biological Action of Ionizing Radiations."

In 1957 the Bibliographical Reference Manual on Iron Ores (edited by Academician Bardin) appeared. It was prepared jointly by VINITI with the cooperation of the Baikov Institute of Metallurgy. This is a capital bibliographic reference work comprising 120 author's quires. It covers over 10,000 titles of Soviet and foreign sources with annotations revealing the subject-matter of the material.

The Institute has been conducting work in the fields of terminology and lexicography. In 1955 two—English-Russian and Russian-English—Dictionaries on Nuclear Physics and Engineering were put out by the Institute for the Atoms-For-Peace Conference held in Geneva that year.

Another important facility enjoyed by our scientists, engineers and technicians is the Institute's photo and microfilm copying service which supplies the former with copies of any original article covered by the Abstracts Journal. In 1957, for instance, the Institute sent out some 400,000 photocopies.

At present about 30 countries subscribe to VINITI publications. The Institute peruses literature in 50 languages arriving from 88 countries. In 1957 in addition to the full list of Soviet literature (pertaining to its particular field) the Institute received 12,250 titles of foreign periodicals.

VINITI exchanges scientific publications with 580 organizations of 46 countries (including Britain, the USA, Canada, France, Italy, Greece, Morocco, India and Japan). The Institute is a member of the International Federation of Documentation and participates in the activities of the ISCU Abstracting Board.

The informational materials are prepared by a large number of highly skilled specialists. Apart from its large permanent staff the Institute recruits the services of numerous supernumerary workers, including full members and corresponding members of the USSR Academy of Sciences, as well as doctors and candidates of sciences.

Prominent scientists exercise direct supervision over the Abstracts Journal series: Academician L.I.Sedov is the editor-in-chief of the “Mechanics” series; N.V.Agheyev—corresponding member of the USSR Academy of Sciences— is the editor-in-chief of the “Metallurgy” series; Academician Artobolevsky is on the editorial board of the “Engineering” series; Academicians L.A.Orbeli and N.V.Tsitsin are on the editorial board of “Biology”; Academicians P.S. Alexandrov and A.N.Kolmogorov are on the editorial board of the “Mathematics” series; and Academician A.F.Ioffe is on the editorial board of the “Physics” series.

Organizationally the Institute consists of editorial offices, scientific departments and research laboratories.

The Institute has eleven editorial offices covering in their Abstracts Journal series the respective sciences (however, the Editorial Office for “Chemistry” also handles the “Biological Chemistry” series and that of “Physics”—the “Geophysics” series). The editorial offices are also responsible for the preparation and issue of the “Advances in Science” reviews in their respective spheres.

The “Express Information” bulletins are prepared by the Department for Scientific and Technical Information.

Among the scientific departments there are also the Acquisition Department and Department for Systematization.

The Acquisition Department has as its province of study the entire world literature on science and technology. It is the responsibility of this department to acquire through subscription and exchange the publications required.

All publications arriving at the Institute are registered and then forwarded to the Department for Systematization.

Here the publications are distributed to the groups organized on the language principle (languages of the USSR and People's Democracies, English, Romanic, German and Scandinavian, and the Oriental languages). The language groups mark the materials for subsequent transfer to the respective editorial offices. For this purpose each article is stamped with the letter index of the editorial office the material is addressed to and the personal number of the marker. It is not infrequent that one and the same article is brought to the attention of more than one editorial office. All patents, however, irrespective of the language of origin, are handled by a special Patents Group which deals with them in similar fashion.

The materials are then forwarded to the Transcription Group which transcribes the names of the authors and firms in letters of the Russian alphabet.

Further the sources are passed on to the Bibliographical Group which prepares bibliographical descriptions. A bibliographical card is filled out for each marked article and the latter is also furnished with a special operational number. Thence the materials are transferred to the Control Group to be checked for factual and spelling errors that may have been made in the bibliographical description.

The publications are next taken to the Technical Processing Group. Here the articles are either photo-reproduced—if the source is available in one copy only —or clipped out —if it is available in two or more copies (one copy of each publication is always reserved intact). Simultaneously, the rotorprinting of the bibliographical cards takes place.

Each article thus processed is provided with a bibliographical card which is attached to it. At the same time cards for the authors' and general catalogues are duplicated. The articles (or photo-copies) are next distributed to the pertinent editorial offices, and the processed periodicals are placed in the custody of the Storage Department.

At the editorial office the articles are delivered to the office bibliographical group which classifies them according to the sections of the particular Abstracts Journal series. The materials are then handed over to the section editor who is in charge of a group of specialist editors organized into a sector. It is the responsibility of the section editor to distribute the materials to the specialists (VINITI'S supernumerary abstractors) and to supervise the editing of the abstracts prepared by them. After the abstracts have been edited they are passed on to the Editorial and Publishing Department which handles the issue of the Abstracts Journal.

It takes from four to six months from the moment the source arrives at the Institute to the time it is reflected in the Abstracts Journal series.

According to their subjects the more important scientific and technical magazines are distributed among the editorial offices on a priority basis and are handled by the scientific departments as soon as they arrive. Each of these magazines shall be regarded as a top-priority objective only by the editorial office concerned. It is the duty of that editorial office to rush the abstracting of all the articles contained in these magazines and to supply the other editorial offices with copies of the abstracts.

The editorial offices may take for their series any article appearing in any such magazine. However, such an article may be re-abstracted only after the abstract received from the priority office has been found unsuitable for the specific purposes of the other editorial offices. Each office decides whether to publish the abstract in its original form, to revise it, or to order a new abstract of the earmarked article.

The Abstracts Journal uses three forms of publications: the abstract, annotation and bibliographical description.

The abstract is the principal form for original works.

The annotation is used for reviews—whether books or articles—and popular science works the subject-matter of which is not fully revealed in their titles, for monographs or reviews either stating new points of view or collating original studies.

The bibliographical description is used for books, reference manuals, textbooks, collections, reviews and popular science articles, non-original articles on materials published earlier and certain articles on adjacent branches of science.

The abstract

The size of the abstract is determined by the value of the article. An average abstract shall not exceed one double-spaced typewritten page. Should the work contain information of considerable value the size of the abstract may be increased to one and a half pages or, in exceptional cases, to two or even three pages.

The abstract invariably gives the object, method, the principal theoretical prerequisites, results of the work, numerical data of scientific-or technological interest, as well as the author's view on the possibilities for applying the results of the work to science and technology. Concrete (numerical) data shall not be substituted by generalities. As a rule the history of the investigated problem is not given.

All non-metric measurements appearing in the article are converted to the metric system when quoted in the abstract.

The annotation

The size of the annotation is determined by the value of the material. Normally it shall not exceed one third of a double-spaced typewritten page. The annotation states as briefly as possible the problems dealt with in the material annotated.

The bibliographical description

The guiding element in the description is the title (not the author) of the work covered because it makes it easier for the reader to look for the publications that may be of interest to him.

If the article is in Russian, the bibliographical description consists of the title, the author's family name, first name (or initials) as well as the account of the source.

If the article is in a foreign language, the bibliographical description consists of, firstly, the Russian translation of the title, and the author's family name in the Russian transcription. Neither the author's first name nor his initials are given in the Russian transcription because (unless the full name is known) it is often impossible to find adequate equivalents for the initials. The title of the article, the author's family name (and his initials) in the original language follow in parentheses. The name of the source (periodical) is abbreviated in accordance with the Institute's list of abbreviations. Then come the publication particulars: year of issue, number of volume (if the volume number is omitted, only the year of issue is given), edition number and language of the original.

In case of pamphlets or books the bibliographical description basically consists of the same elements: title, sub-title, number of volume or part and, if available, its heading, edition number, name of author, place of publication, name of publishing house, year of issue and circulation figures. As to foreign publications all information is given in the language of the original in parentheses; only the name of the book is given in the Russian, and the author's name in the Russian transcription.

In case of authors' abstracts of dissertations the bibliographical description reveals the subject of the dissertation, the author's name, the degree he is seeking, place where the dissertation is to be defended, place and year of publication of the author's abstract.

In case of reviews the bibliographical description consists of two parts: one—account of the work under review giving all its publication particulars in the above manner, and two—specification of the source carrying the review in question. The family name of the author of the review and the name of the source carrying the review are enclosed in square brackets. The title of the review is not quoted.

A patent description includes: the name of the invention, the name of the inventor, the owner of the patent (name of individual, firm or government office), country of origin, classification index, patent number and date of publication.

These elements are given in the following order: name of the invention in the Russian, family name of the inventor in the Russian transcription; then in parentheses in the original language—the name of the invention, family name of the inventor, his first name or initials; separately in square brackets (also in the original language) the name of the owner of the patent (if the owner is not the same man). Further comes the abbreviation “Pat.” followed by the country of origin, classification index, patent number, date of publication and source containing information on the patent (unless the source is an official patent journal).

The title of the article is given in full as in the source. If the title is given in several languages the description uses the language of the article.

Today an information agency is unable to operate efficiently unless it enjoys the advantages of electronic computers and other modern machines. The value of information is not only in its completeness but also in its expediency.

VINITI has two laboratories working on problems aimed at mechanizing the information service.

The Laboratory for Mechanization of Information Work has arrived at a stage when it is able to launch an experimental search for information in the field of mechanics. Apart from that machines have been used for compiling an experimental copy of an authors' index for the “Electrotechnics” series.

The Electric Modelling Laboratory has been assigned the following tasks: elaboration of the theory and principles for the designing of information machines; development of the theory of new methods of electric modelling; creation of new technical means for modelling and design of information machines. In 1957 an experimental machine consisting of magnetic elements, ferrite and capacity memory units was developed in the main. The machine is programmatically controlled.

To sum up: This account is intended as an insight into the operational mechanics of the All-Union Institute for Scientific and Technical Information. In our opinion, the approach to the problem of total information adopted by the Presidium of the USSR Academy of Sciences in 1952, has been fully warranted by the Institute's experience of five and a half years. It has confirmed that only a centralized information service is capable of yielding exhaustive information on the advance of world scientific and technological thought.

A task which may be insurmountable for an individual specialist can be coped with successfully by a highly-skilled multifarious scientific body. As we see it, our object is to make the Abstracts Journal into a sort of a scientific forum for intellectual exchange between the scientists of all countries.

Naturally, we have not yet solved all the problems confronting us. Vast fields of science and technology—medicine, agriculture, construction and transport—have yet to be covered by the Abstracts Journal. We have our shortcomings. The gravest in our view is the considerable delay in the issue of indexes.

We hope that our paper will evoke helpful comment and criticism aimed at improving our work, promoting friendly relations and mutual understanding.

SUMMARY OF DISCUSSION

In introducing the Area 2 panel discussion, Dr. Elmer Hutchisson first defined an ideal abstracting service as one which would cover all articles in its own and borderline fields, would condense to infinitesimal length the significant contents of these articles, would provide unambiguous abstracts and indexes in every reader's own language with zero (or negative) time lag, and would cost nothing. Recognizing that such specifications are not likely to be met, he pointed out that compromises must be made and said the panel had decided to organize the discussion around two basic functions of an abstracting service, with each member speaking briefly on a particular aspect or problem in relation to one of these functions. The functions and sub-topics were as follows:

Function 1: To assist scientists in keeping abreast of scientific progress

A. Promptness of publication

B. Quality of abstracts

C. Per cent of condensation

D. Classification

E. Subject coverage

F. Economy in abstracting

Function 2: To assist scientists in retrieving specific information

A. Coverage for retrieval

B. Indexing

C. Economy of space

D. Monetary economy

The Chairman closed his introductory remarks by reminding the audience that panel members are likely to be more difficult than subject matter to organize and predicting that these panel members would use considerable freedom in deciding what to discuss under the assigned topics. Summarized below are each panel member's remarks and those of the registrants who spoke from the floor during the discussion periods.

PROMPTNESS OF PUBLICATION

The Chairman of the Panel discussed the first sub-topic. To point up the problem of promptness in abstracting, Dr. Hutchisson presented a scatter diagram of representative data on time lags *(a)* from receipt of a paper by an editor to its publication and *(b)* from publication to appearance of the abstract. The former varied from two or three months to a year to two; the latter, if a new abstract is written, from three to six months. He contrasted these delays with the hour or two required by a newspaper, or the seconds needed by television, to report the latest murder.

He then mentioned three current approaches to improving promptness. One is to airmail proof to abstract journals; this is being done in the physics field and is discussed in the paper by Dr. B.M.Crowther. A second is exemplified in the new *Physical Review Letters* in which its editor, Dr. S.A.Goudsmit, attempts to report recent research within two weeks; in fact, by including some abstracts of papers still to appear in the *Physical Review,* he is achieving a measure of negative time lag. The third approach cited by Dr. Hutchisson is the use by abstracting services of the so-called "author" abstract—that is, the abstract which appears with a paper when it is first published. He noted that this third method raises a controversial point of long standing and one about which much more undoubtedly would be heard during the afternoon.

QUALITY OF ABSTRACTS

Dr. S.H.Gould limited his remarks on this sub-topic to one phase of quality in abstracting, that of subject slanting as discussed in some detail in the paper in which Mr. Saul Herner reports on his analysis of abstracts of the same paper appearing in two or more of nine major abstracting journals. For the more than 200 papers studied, no significant slanting occurred toward the particular subject field of the abstracting journal and, in fact, seldom did the abstracting journal's product differ appreciably from the original abstract. Dr. Gould said he found considerable encouragement in these results because he frequently "borrows" abstracts from other journals for use in *Mathematics Reviews*. He noted that this practice might itself appear to introduce considerable delay, since he ordinarily does not obtain such an abstract until it has appeared in its regular journal. Actually it does not do so, however, because he indulges in such borrowing only after he has been unable to get the abstract from one of his own abstractors. He felt that the results of the Herner study indicated that multiple use of good author abstracts will not seriously reduce abstracting quality.

DISCUSSION FROM THE FLOOR

Sir Herbert Howard referred first to statements made in the morning session to the effect that scientists do not make a great deal of use of abstracts. He then said that his organization (Commonwealth Agriculture Bureau) issues 14 abstracting services, the production of any one of which requires perusal of several thousand serials in some 30 languages. Since he believes no single scientist, working entirely on his own, could possibly keep abreast of this much material, he suggested that the morning statements indicated either bad abstracts or bad scientists.

Apropos of Sir Lindor Brown's emphasis the previous evening on the importance of liaison between the producer, the distributor, and the user of scientific information, Sir Herbert described the manner in which this problem is handled in his organization. Each editor of an abstracting journal, and as far as possible the abstractors also, must be versed in its subject field. Each individual bureau is located in a research institute dealing with its subject. Each journal is aimed at a particular set of readers and liaison is encouraged between scientists, editors, and bureau directors. Further, official correspondents, also familiar with the particular subject field, are established in the various countries of the Commonwealth.

Dr. J.E.L.Farradane, referring to the Herner paper mentioned by Dr. Gould, said he believes that there should be some degree of subject slanting in a specialized abstracting journal and that this is why readers go to a particular journal. Therefore, if the Herner findings are correct, he thinks something is wrong with abstracting. Also, he thinks that one should not think of an author abstract as being unslanted. Dr. Farradane believes instead that it usually is slanted in what seems to him to be the worst possible way—toward what the author hoped to do and thinks he wrote, but not necessarily toward what he did do and write. Therefore, in Dr. Farradane's opinion, a specialized abstracting service never should lift an abstract directly from the original journal.

PERCENT OF CONDENSATION

Professor J.D.Bernal began his comments on this sub-topic by saying he would make his per cent of condensation as large as possible. He then mentioned two aspects of “keeping up” with science—one, staying abreast on a contemporary basis and the other, catching up. He suggested that the time lag in issuance of abstracts may equal or exceed the “half life” of the material abstracted; in which case abstracts can be of little use for the former function. For the latter aspect, however, they can be very valuable to researchers starting new projects, to teachers, to writers, and the like. He believes the typical scientist uses abstracts to bring his knowledge within about three years of the present and then goes on his own, searching journals, talking to and corresponding with scientists, and so forth. For this kind of use, Professor Bernal finds the present abstracting pattern fairly adequate.

Regarding scope, he would argue for the indicative (or descriptive) type of abstract in general journals and the informative type in special journals with the latter always being subject slanted. He said that in the special abstract journals with which he is familiar, the abstracts are slanted. Finally, Professor Bernal raised the question of whether abstracts are needed at all. Referring to the papers by Dr. Isaac Welt and Mr. Eugene Garfield, he suggested that the need is not for lists of publications but for well-indexed lists of facts. He believes such indexes might well take the place of indicative abstracts while providing an important supplement to informative abstracts. In closing, Professor Bernal suggested that careful consideration be given to varying the length of the abstract with the importance of the paper.

DISCUSSION FROM THE FLOOR

Dr. Chauncey D.Leake strongly urged condensation of abstracts through use of a telegraphic style, saying he believes at least 50 per cent reduction in length could be achieved with no loss in reader comprehension.

Dr. S.R.Ranganathan urged publication of abstracts in a strictly and minutely classified subject sequence. He believes abstracts then could be limited to new facts and ideas with much of the material that now appears in them being taken care of by the subject headings.

Dr. Farradane said, “Heaven save us from the telegraphic abstract.” His experience as editor of an abstract journal indicates that people do not get adequate information from such abstracts and he strongly urges complete sentences, logical sequence, and readable style.

Dr. John W.Tukey suggested that for most people keeping abreast means trying to be *(1)* fully informed on the literature in a field of zero width and *(2)* progressively less well informed as the field grows wider. He proposed that papers be given a “surprise” rating with abstracts of those having a high rating appearing in general journals and those with zero surprise being published only for narrow specialists.

Dr. Hutchisson pointed out that very few of the papers in the ICSI preprint volume carry abstracts.

CLASSIFICATION

Dr. George Shortley introduced the sub-topic of classification. He noted that much of the material involving storage and retrieval in papers of Areas 4, 5, and 6 is relevant to the classification problem. He then raised the question of what role the major abstracting services should play in the classified organization of scientific information. He asked, “Who better than the abstractor can classify an article according to any given scheme or select the key words for any system of correlative or coordinate indexing?" He suggested a pattern whereby abstractors would supply such information along with the abstracts, with the key words then providing the basis for clerical selection of papers and abstracts to be indexed by specialized laboratories and libraries, or to be routed automatically to individual scientists. In closing, Dr. Shortley referred to an Area 5 paper by Dr. E.J.Crane and Dr. C.L.Bernier which points out that multitudes of users of present documentation are in the future, many yet unborn. He then asked what plans the editors of today's abstracting services are making, looking toward systems of information retrieval that will meet the needs of the future.

DISCUSSION FROM THE FLOOR

Dr. Crowther stated it to be his very strong feeling that a wide classification scheme suitable for use by a major abstracting journal can serve as only the first step in any useful retrieval system. Individual users have a vast variety of specialized needs which the abstracting service is unaware of and therefore cannot meet.

Dr. Ranganathan disagreed strongly with the idea that the abstractor is the best classifier, suggesting that each has a highly specialized task and should be a specialist in it.

Three speakers dealt briefly with the matter of education in the use of scientific literature. Mr. Gerald Estrin urged that greater emphasis upon such education would be one of the most useful ways of helping the scientists of the future. Sir Alfred Egerton mentioned that this was one of the recommendations of the 1948 Royal Society Conference in London. Mr. Herner cautioned against overemphasis on uniformity in education for the use of literature lest the needs of the scientist be subordinated to the vagaries of the systems and "1984" arrive even before 1984.

SUBJECT COVERAGE

Mr. G.Miles Conrad began the discussion of this phase of abstracting and indexing. He noted first that when users are asked what is most important in the abstracting field they invariably reply, "Complete, prompt coverage, and prompt indexing." He then spoke briefly about five phases of coverage. First, he pointed out that any statement regarding percentage of coverage by an abstracting service implies knowledge of the total size of the literature in that field. Only if one knows what exists can he determine whether or not adequate coverage is being achieved. Here, he complimented Miss Estelle Brodman and

Mr. Seymour Taine on their paper, a quantitative survey of medical literature, and urged that similar studies be made in other fields. Knowing what literature exists, one then can start trying to acquire it—his second point. Mr. Conrad stated that conventional periodicals present no acquisition problem. Where language or obscure origin poses difficulties, interservice cooperation can be employed effectively along lines described for exchange domestically in the paper by Dr. Otto Frank, internationally in Dr. Crowther's article, and within a subject field in Dr. A.B.Agard Evans' discussion of building literature. An acquisition problem of increasing seriousness is presented by unpublished documents such as conference papers and reports of various kinds. A paper by Mr. Felix Liebesny suggests that a large proportion of the former is destined to be lost. Mr. Smith's paper discusses the latter type of literature at some length as to nature, scope, and availability.

Third, Mr. Conrad pointed out the increasing importance of the cost factor as complete coverage is approached. Here he found particularly pertinent the cost analyses for different levels of coverage presented in the paper by Dr. and Mrs. Malcolm Rigby. There is, of course, the question of whether complete coverage is desirable—his fourth point. Here he cited two papers which seemed to a degree to debate this matter. In one, Dr. C.S.Sabel, after reviewing numerous unclassified reports, stated, “The references in an incomplete list of documents are unlikely to indicate more than a small proportion of the remaining documents.” However, the Lykoudis, Liley, Touloukian paper suggests as an adequate approach, obtaining leads from abstracting services and then approaching bibliographic completeness from the references in the primary articles so located.

In considering his fifth point, promptness, Mr. Conrad referred to the Herner paper on subject slanting mentioned above. He emphasized the great saving in time that can be achieved by using author abstracts and suggested that these can be made acceptable if authors are required to submit them with the manuscript and if journal editors take editorial responsibility for them. He noted that the ICSU Abstracting Board has urged primary journals to follow this procedure.

In closing, Mr. Conrad emphasized that deficiences in coverage by abstracting services almost without exception are due to insufficient resources of personnel and funds and not to indifference to the growing literature.

DISCUSSION FROM THE FLOOR

Dr. Hutchisson suggested that someone might wish to comment on whether abstracting services should cover some journals completely and others partially, or only the important papers in a large number of journals.

On this point, Dr. Farradane commented that type of coverage is a function of whether or not the abstracting journal claims to be complete. He then went on to suggest the danger of too much coverage—that perhaps the larger abstracting journals are becoming so large as to be almost unreadable and that maybe the day of the comprehensive abstracting service is over. He suggested that the reader might be better served by smaller, special-field services which cover borderline material reasonably accurately and adequately.

Dr. Crane commented that whether all or part of a journal should be covered is wholly a function of subject matter. Regarding Dr. Farradane's suggestion, he pointed out the overlapping that necessarily would occur if, say, 20 abstracting journals were replaced by 100 highly specialized ones. He recognized that some journals are getting very large and the time may come when division is necessary but said a single journal still is the most economical way to serve the many people working in a field as broad as chemistry.

Dr. Russell Ackoff cited the following data from a Case operations research study: That in 1957 the average industrial chemist read 0.5 per cent of the articles abstracted in *Chemical Abstracts* and about 1.5 per cent of the abstracts themselves; and that if he had devoted all his reading time to abstracts, he would have read 13 per cent of them.

Sir Herbert Howard spoke briefly on author abstracts. In the 14 abstracting services issued by the Commonwealth Agriculture Bureau they occasionally use author abstracts but mostly find them unsuitable. The principal reasons he gave for this are that good scientists frequently are poor abstractors and sometimes even put things in an abstract that are not in the parent paper. He stated that in his journals the abstracts definitely are subject slanted with three or four differing abstracts of the same paper frequently being prepared for different journals.

Dr. Leake commented that in his opinion abstracting services do a very good job on the facts published in their fields, but to a considerable extent ignore the concepts which are so important to scientific progress.

ECONOMY IN ABSTRACTING

Professor G.A.Boutry began this part of the discussion by observing that the extensive and intense criticism one hears of abstracts at least indicates they are widely used. He suggested that it may be unrealistic to expect both a perfect abstract and high speed of issuance. Any processing of information introduces some deterioration in quality and, other things equal, the greater the speed the greater the loss in quality. He then traced briefly the history of abstracting saying that it originated as critical reviewing which, for the most part, evolved into non-critical abstracting because of the shortage of competent reviewers, that is, reviewers who know as much or more about a subject than do the authors of the papers in the field. He admitted the validity of the criticism that author abstracts often are not very good, and then asked why this is so. His own answer was simply that authors in general lack education in writing abstracts. He referred to the rules for the preparation of synopses (or abstracts) which came out of the Royal Society conference and re-emphasized a point made earlier—that author abstracts might become quite satisfactory if editors of primary journals would promulgate such rules and then take as much editorial responsibility for the abstracts' adequacy as they do for that of the papers.

DISCUSSION FROM THE FLOOR

Supporting Professor Boutry's remarks, Mr. Joseph Hilsenrath suggested that if a scientist is competent to judge and abstract other scientists' papers, it should be possible to train him to abstract his own. He then directed a question on this point to Dr. Crane.

In reply, Dr. Crane agreed that such might be true for many scientists but doubted whether such training could be adequately given to, say, 200,000 authors of papers. He went on to say he believes an author abstract should accompany every published paper. *Chemical Abstracts* uses these as a starting point and then adds whatever further information is necessary for satisfactory indexing. He pointed out that author abstracts mostly should not be this inclusive. For example, if a paper mentions 100 compounds, obviously the accompanying abstract in the journal should not also list all of them. If, however, they do not all appear in the *Chemical Abstracts* abstract, the material will not be adequately indexed and the bibliographic record of chemistry will not be complete.

Dr. Ranganathan commented that it is an over-simplification to assume that an author who can abstract other scientists' articles is competent to abstract his own. He suggested an analogy here with the doctor who diagnoses the illnesses of others but does not try to diagnose those of himself or his children.

Dr. Crowther observed that much misunderstanding is caused by the use of the term “author abstract” and reminded the group that the resolutions of the 1948 Royal Society conference referred to summaries which accompany papers and urged editors to take responsibility for the quality of these. As a case in point, he spoke of one U.S. physics journal whose “summaries” he finds almost wholly satisfactory for *Physics Abstracts* and whose editor does take particular care to see that they are adequate.

Dr. Bentley Glass stated that the U.S. Conference of Biological Editors has gone on record recommending that an abstract be considered an integral part of every published paper and that the journal editors take as much care with the abstract as with the manuscript itself.

Commenting on Dr. Glass's statement, Dr. Tukey suggested that the recommendation of the biological editors does not go far enough. It implies simply an editorial responsibility with respect to the author-written abstract; he believes the basic responsibility to provide good abstracts should lie directly with the editors. Also he would prefer the term “journal abstract” to either “author abstract” or “summary.”

Mr. Paul Lykoudis suggested that there may be disadvantages in too nearly perfect abstracting. He would like to see surveys made *(a)* of all the ideas that have been lost because comprehensive abstracting seemed to tell everything and so killed initiative, and *(b)* of all the unnecessary research that has been performed because of the exhaustive abstracting of papers containing misinformation. He favors some specific fraction of failure in efforts of our good abstracting organization.

COVERAGE FOR RETRIEVAL

In introducing this sub-topic, Dr. Maurice B.Visscher first referred to two Area 1 papers. One, by Dr. Bentley Glass and Mrs. Sharon H.Norwood, indicates that only about 5 per cent of the scientists they interviewed learn about work crucial to their research by reading abstracts. The other, by Mr. R.M.Fishenden, placed the figure at 6 per cent for pure research scientists but at some 15 per cent for applied researchers. Dr. Visscher suggested, first, that one should not infer from the basic research percentages that abstracts are not worth bothering with. Personally, he finds he uses abstracts for two principal purposes—to lead him to very specific data and to provide him with general information in scientific fields other than those of his own research. The fact that abstracts may be the source of a relatively small fraction of new ideas directly pertinent to his personal research in no sense means they have not been immensely valuable to him as a working scientist.

Apropos of the Fishenden findings, Dr. Visscher said one might expect different kinds of scientists and researchers to use abstracts to varying degrees. In this connection, he cited the Smith paper which reports on a series of 50 bibliographic searches and states that not nearly all of the information that was found could have been located if numerous abstracting and indexing journals had not been available.

Next, on the question of the period of usefulness of scientific information, Dr. Visscher cautioned against being too strongly influenced by data cited earlier indicating that the “half life” of scientific research knowledge is very short. He suggested that this varies greatly with subject field and cited examples from biochemistry, biology, genetics, and pharmacology of papers published 40 or more years ago which still are extremely useful.

He said the fundamental question is not whether we need abstracts and indexes—he believes it is obvious that we do. Rather, he thinks the question is what degree of extensiveness and completeness of coverage is necessary, reasonable, and feasible. Here he pointed out the importance of distinguishing carefully between specialized and general abstracting services. To be most useful the former should be complete for the particular area dealt with and function served. The latter cannot be complete and, indeed, should not try to be lest they become so large and unwieldly as to lose their usefulness. In closing, Dr. Visscher urged that economic considerations not be permitted to be limiting factors in provision of good abstracting. He believes that after hundreds of millions of dollars have been spent to obtain research information, no rational society will permit a small fraction of this amount to stand in the way of dissemination of the knowledge.

DISCUSSION FROM THE FLOOR

Mr. R.A.Fairthorne said it was his impression that a special abstract was being thought of as one which attempts to express the relevance of a paper to a certain field, and a general one as a description or condensation of a paper's contents. He suggested that the former objective is relatively hopeless of achievement, since it implies a kind of omniscience on the part of the abstractor. He urged that abstracting efforts be concentrated on the general type.

Professor Pierre Brygoo said he sees two basic uses for abstracts: *(1)* to acquaint a reader with the contents of an article without him having to read it; and *(2)* to steer him to articles which he should read. He questioned whether present informative abstracts fulfill the former function very well and suggested that the problems involved in doing so are much more formidable than is commonly recognized. The ratio between reading times for an article and its abstract is not necessarily the same as that between their lengths. He has found occasionally that reading the abstract even required an appreciably longer time than did perusal of the entire paper. He then made two suggestions: *(1)* That if abstracts serve only to direct users to papers they should read, the indicative type, which is both cheaper and faster to prepare, should suffice (except where articles may be impossible or very difficult to obtain); and *(2)* if actual summaries of contents are to be provided, the informative abstract is not the best way to organize the material.

Dr. Hutchisson referred back to previous remarks on the "half life" usefulness of research data and pointed out that this must vary greatly depending upon whether one is thinking of researchers directly in the field concerned or of those working in borderline areas.

INDEXING

Dr. Crane restricted his remarks on this phase to subject indexing, regarding which he stated the following principles: *(1)* that indexing is fully as big and important a job as abstracting; *(2)* that indexing frequently has been seriously neglected in the planning of information services; *(3)* that good indexing is based on subjects, not on words; *(4)* that good indexing is not classification. The *Chemical Abstracts* staff devotes half of its over-all effort to indexing and finds that from two to five years are required to train a good indexer, starting with a Ph.D. in chemistry. They believe only very complete indexing can insure adequate retrieval. *CA* indexing averages at least six index entries per abstract.

In closing, Dr. Crane said he recognizes the need for and importance of mechanization in indexing; however, he cautioned that if speed is gained at the sacrifice of quality, the final bargain will be a bad one.

ECONOMY OF SPACE

Dr. Williamina Himwich introduced this sub-topic by summarizing some of the advantages and disadvantages of microfilm and similar space-saving developments and suggested that perhaps the principal reason problems in this area have not been more completely solved is that the space situation has not yet become really desperate. She outlined the advantages of distributing abstracts on cards which permit the recipient to weed out the “duplicates” received from different services, as well as cards of no interest to him, thereby permitting him to retain in a relatively small space a collection of useful abstracts. She pointed out that such a system allows the retention of peripheral field items with a minimum of difficulty; she then went on to discuss the importance of such material.

She suggested that enabling a scientist to see and review information in fields other than that of his immediate research concern may be the most important function of abstracting and indexing services. In his own narrow-area of research the scientist knows the principal workers, attends meetings, corresponds with colleagues and may be able to “keep up” quite satisfactorily without much outside help. It is in borderline areas of possible, but not always obvious, significance to him that he particularly needs bibliographic assistance. Here interpretation of information bearing labels other than that of his personal research is all important; abstracts and indexes can bring this material to his attention.

MONETARY ECONOMY

Dr. Allen J.Sprow in discussing monetary economy first pointed out that one cannot realistically discuss retrieval of scientific information without bringing in monetary economy and that, whether one likes it or not, the question "For how much?" often has to take precedence over "For what?" and "For whom?" Apropos of this statement he cited the large fraction of Area 2 papers which deal in some degree with cost—either primarily or in part. A number of these discuss procedures in which greater economy is either a principal objective or a by-product. Such approaches include cooperation and coordination among services as considered in the papers of Drs. Crowther, Evans, and Frank and national centralization as dealt with in Professor Mikhailov's paper. Dr. Sprow then went on to discuss briefly the recently formed National Federation of Abstracting and Indexing Services, the mission of which is to improve the abstracting, indexing, and analyzing of the world's scientific and technological literature in such a manner as to make it readily and promptly available to the scientific community. The Federation hopes to achieve this objective through cooperation among its members in exchanging their several outputs, in conducting mutually-beneficial studies, in discussing common problems, and in jointly preparing various bibliographic tools needed by all.

DISCUSSION FROM THE FLOOR

Dr. Jesse H.Shera mentioned two fallacies which in his opinion had permeated the afternoon's discussion. The first, he said, was the assumption that an abstracting service can be all things to all men. He emphasized that this is not true—pointing out, for example, that abstracts are excellent for browsing but may be useless for retrospective searching, while a fully mechanized system may greatly facilitate searching but "you can't do much browsing in a magnetic tape." He suggested that some abstracting services have gone far beyond the point of diminishing returns in their attempts to do everything for everybody.

He said the second fallacy was more difficult to define but concerned the relationship of the whole to the sum of its parts. An abstract is a unit but in general we consult abstracts as a conglomeration of many such units which, as a "whole," acquire characteristics that are not simple sums of the traits of individual abstracts.

He felt the Conference had been trying to reconcile things which are basically irreconcilable and suggested that before solutions to the problems of abstracting can even be approached, firm and realistic definitions must be established as to what abstracting services can do best. Then efforts should be concentrated to these ends.

Dr. Farradane returned to the question of indexing and suggested that if indexing requires half of the effort of an abstracting service, this is an area in which research definitely is needed. He then took issue with Dr. Crane's earlier statement that indexing is not classification. He made reference to faceted classification work of his own in which subject indexing is carried out purely by classification principles. He believes this approach achieves a much greater uniformity of style in a subject index than ordinarily is the case.

In reply, Dr. Crane reiterated his previous statement, saying, however, that to some extent indexing uses classification methods—that is, that classification is just one of the indexer's tools.

Mr. J.J.O'Connor cited a statement in the paper of Lykoudis, Liley and Touloukian to the effect that in their work a cover-to-cover search had to be made of *Chemical Abstracts*. He asked whether some notification of the *CA* subject index might have made this unnecessary. Dr. Touloukian replied that while for their particular use, *CA* does not give them full coverage of what they require, they recognize it as being a most thorough journal which adequately serves the purpose for which it was designed.

Mr. Garfield, an Area 2 author, said that since no one else had criticized his paper on citation indexes he wished briefly to do so. He then mentioned two weaknesses in the universal applicability of such indexes—both stemming from the possible selection of references by the author on bases other than strict scientific pertinence. One is the possible desire on the part of an author to protect his own priority in the publication of certain research results. The other concerns the desire a scientist might have not to jeopardize his avenues for obtaining grant funds to support his research. Mr. Garfield suggested that foundations and other agencies providing research support should perform more thorough literature searches before making grants.

In closing the session, Dr. Hutchisson expressed his appreciation and thanks to the authors of papers in Area 2, to the panel, and to those who participated from the floor in the discussion of the function and effectiveness of abstracting and indexing services.

DWIGHT E.GRAY, *Rapporteur and Area Program Chairman*
ELMER HUTCHISSON, *Discussion Panel Chairman*

AREA 3

EFFECTIVENESS OF MONOGRAPHS, COMPENDIA, AND SPECIALIZED CENTERS

Present trends and new and
proposed techniques and types of services

AREA ORGANIZATION

Authors of Papers

Members of Discussion Panel

Chairman: ALEXANDER KING, European Productivity Agency, Paris, France

DALE B.BAKER, The Chemical Abstracts Service, The Ohio State University, Columbus, Ohio

PIERRE BRYGOO, Scientific Documentation, Institut Pasteur, Paris, France

RALPH CLARK, Stanford Research Institute, Washington, D.C.

BERNARD M.FRY, Office of Scientific Information, National Science Foundation, Washington, D.C.

JOHN C.GREEN, Office of Technical Services, U.S. Department of Commerce, Washington, D.C.

MASAO KOTANI, Science Council of Japan, Ueno Park, Tokyo, Japan

Z.MAJEWSKI, Central Institute of Scientific and Technical Documentation, Warsaw, Poland

A.I.MIKHAILOV, All-Union Institute of Scientific and Technical Information, Academy of Sciences of the U.S.S.R., Moscow, Russia

E.PIETSCH, Gmelin Institute, Frankfurt/Main, Germany

PROPOSED SCOPE OF AREA 3

IN THIS AREA, perhaps more than in any other, one finds a great disparity in the scientific information services and facilities that are available in various disciplines. Many of these differences result from basically different characteristics and problems of the various sciences. Others, however, seem to stem from a lack of organization, experience, or perhaps awareness of the feasibility of adequate bibliographic control.[1]

In some areas of science, reviews are in their infancy. In medicine, however, the volume and importance of reviews have been sufficient to stimulate the preparation of a *Bibliography of Medical Reviews.* Some reviews are little more than indexed or annotated bibliographies, whereas others provide more detailed treatment of narrow fields, and a few, such as *Nutrition Reviews,* attempt a highly critical appraisal of recent work, both individually and collectively, with a view to influencing the direction of future research.

In the matter of compendia there is also a marked difference between fields. A brief search of the chemistry collection at the National Bureau of Standards Library revealed approximately 50 compendia. Yet a search of compendia, similarly defined, in physics yielded barely a dozen titles.

Increased research activity has meant not only increased publication but also more activity in establishing specialized centers for the collection, correlation, and dissemination of scientific information. The classic compendia are now out of date and it is doubtful indeed whether their revision, even on a much augmented scale, will meet future or even current needs. The Landolt Börnstein tables are now being revised in approximately the same format, but the Gmelin Institute is undergoing a significant reorganization—away from compendia and toward reference service. Compendia and reviews have played an important part in the conduct of scientific work and will undoubtedly continue to do so. In many specialized areas, nevertheless, more expeditious collection and more prompt and direct dissemination are required. Here specialized centers or organizations have been established. Some examples of such United States centers are:

The Chemical-Biological Coordination Center

[1] The proposed scope of the Conference Area, as shown here, was prepared during the Spring and Summer of 1956 and provided to all potential contributors as a guide to the aims of the Conference.

Snow-Ice and Permafrost Research Establishment

Bio-Sciences Information Exchange

American Petroleum Institute Project 44

Manufacturing Chemists Association Project

Committee on Spectral Absorption Data

Nuclear Data Project

Solid Propellants Information Agency

Liquid Propellants Information Agency

Bibliography Section T.I.D., Library of Congress

Thermophysical Properties Research Center, Purdue University

Human Relations Area File

Historical Records Project, George Washington University

Linguistic Center, Indiana University

Collection of Survey Materials, Columbia University

A number of European centers are interesting in this connection. Some examples are:

Gmelin Institute

Centre de Documentation, CNRS, Paris

Centre de Documentation, IRSID, Saint-Germain

Dokumentation der Molekul-Spektroskopie, Weinheim

All-Union Institute of Scientific and Technical Information, Academy of Sciences, USSR

It is to be expected that such centers will be the first to apply the latest techniques in rapid collection, search, coordination, and dissemination. It is also probable that these centers will expand their services as these new techniques become available. While much of the change may be in the direction of increased and streamlined reference services, a significant increase is to be expected in published abstract bulletins, bibliographies, and technical compendia.

It will be important, therefore, to discuss fully the organizational, economic, and technical characteristics of these specialized information centers with particular reference to the impact of machine technology on the organization and services of such centers.

This working area should be directed to the following problems:

1 Summary of studies already made of book publication patterns, existing compendia in the major branches of science, and features of existing specialized documentation centers.
2 Evaluation of the effectiveness of a number of typical specialized services.
3 Assessment of these services in terms of new and proposed techniques (such as mechanized searching systems) and types of services that might be established.

SUGGESTED CONFERENCE PAPERS

1 Characteristics and effectiveness of the monographic literature: A review of the role of scholarly monographs in the utilization of scientific information. The paper should include a discussion based on studies of the way scientists in various fields depend on the monographic literature as contrasted with periodicals.

2 Scientific reviews in the natural sciences: A study and analysis of the development, trends, objectives, and effectiveness of the review literature in the various natural sciences.

3 Reviews in the medical sciences: See item 2 above. The *Bibliography of Medical Reviews* should provide a good starting point for the analysis.

4 Survey and analysis of existing compendia in the physical sciences: The *Index of Mathematical Tables* by Fletcher, Miller, and Rosenhead and the quarterly *Mathematical Tables and Other Aids to Computation* which supplements it provide an excellent example of what can be done to keep track of an ever-growing collection of data. The need in the physical sciences has been recognized for some time.

5 Survey and analysis of existing compendia in the medical and biological sciences.

6 Papers outlining in considerable detail the objectives, organization, staff, facilities, operation, financing, services, effectiveness, and future plans of various information centers.

7 A directory and an analytical summary of services provided by specialized information centers in the United States.

8 A directory and analytical summary of services provided by specialized information centers abroad.

Review Literature and the Chemist

DENNIS A.BRUNNING

ABSTRACT. A questionnaire was circulated to 150 scientists engaged in various fields of chemical research. This paper is based on the 65 completed questionnaires which were returned and represents an attempt to discover how far the various types of scientific review are necessary to the chemist engaged in academic or industrial research and to ascertain how he makes use of the review literature and whether it is adequate for his needs.

The great volume of literature reporting original work has grown steadily year by year and has, under present methods of publication and information dissemination, become almost unmanageable. There has been a corresponding increase in the size of abstracting journals and an increase in the number of review articles and review journals, but these publications, which may be regarded in part as keys to the literature, are beginning to add to the difficulties of information retrieval rather than to facilitate it. The large volume of research currently being undertaken has led to a corresponding demand for more critical interim summaries of work in progress and to periodic comprehensive surveys of particular fields. The complexity of modern science has also enforced an increasing specialisation upon the individual scientist who at the same time is continually finding that his specialisation impinges upon others and thus makes it imperative for him to keep abreast of advances made in fields adjacent to his own. It is the review article and review journal which attempt to satisfy these needs by accumulating, digesting, and correlating the current literature in particular fields and giving an indication of the direction which future research might take. The review literature is therefore coming to play an ever increasing part in the dissemination of the results of research, and it is becoming more and more essential that this form of literature should be organised rationally and should be closely geared to the needs of the scientist. It is well to remember

DENNIS A.BRUNNING Chester Beatty Research Institute, Institute of Cancer Research, Royal Cancer Hospital, London.

that at this time the scientist has more literature to read and less time in which to read it than ever before. In conversation with scientists one continually hears complaints about the amount of literature being produced and the methods of storage and retrieval, and it was therefore surprising to the writer that in the answers to the questionnaire the general feeling seemed to be one of satisfaction so far as review publication was concerned.

THE QUESTIONNAIRE: DISCUSSION OF ANSWERS RECEIVED

The majority of chemists read reviews in their own and related fields, but they read reviews of a general scientific nature to a much less extent. Generally speaking there are enough reviews being published and very few chemists feel that there are too many. The *Advances in*—and *Annual Reviews of*—series are found useful, but it is a common complaint that they are out of date on publication and that they tend to become simply a collection of unrelated summaries and sometimes degenerate into a mere list of references. Many of the comments made, as one might expect, tend to cancel out one another, usually because of the differing viewpoints of those making comments. Chemists engaged entirely on research need comprehensive reviews on specialised topics supplemented with full and complete bibliographies, whilst those concerned mainly in lecturing and teaching ask for broadly based reviews with key references. Several chemists suggested that the cost of these annual review volumes was too high for individual purchase and a scheme whereby individual review articles could be purchased as offprints would be very useful. This is not a new suggestion, for it was advocated for original papers by J.D.Bernal at the Royal Society's Conference in 1948, albeit without obtaining much support. It is suggested that reviews form a section of the literature which would be a very suitable one for an experiment in the distribution of scientific information along those lines.

One of the main difficulties in obtaining good reviews is simply that not enough qualified people are prepared or able to give the time necessary to produce them. There are still far too many “hack reviews” being published, and one frequently sees more than one review on the same subject by the same author which tends largely to be a catalogue of the author's own work. It is often a fault of editorial policy that too little space is allowed to a reviewer for him to do justice to his subject. No one wishes to have reviews of unnecessary length, but if too little space is given, then authors will naturally concentrate on those aspects of the subject under review which appeal to them and ignore other lines of research. Undue compression will also make the review incomprehensible except to the specialist.

The most frequently read review journals are *Annual Reports on the Progress of Chemistry, Quarterly Reviews of the Chemical Society,* and *Chemical Reviews* (A.C.S.). Other reviews specifically mentioned were the special review issues of *Analytical Chemistry and Industrial and Engineering Chemistry* and the *Advances in*—volumes which were relevant to each chemist's area of specialisation. Several industrial chemists considered that much good review material was to be found in the private or semi-private house journals of industrial organisations and in general those working in the industrial field are less enthusiastic about the *Annual Reviews of*—and *Advances in*—series. (A complete list of review journals mentioned appears in Appendix 3 to this paper).

There is practically unanimity of opinion as to what qualities go to make a good review. The criteria most quoted as desirable were: (1) expert writer, (2) critical approach, (3) comprehensive, (4) clarity and balance, (5) good bibliography, (6) synopsis —use of tables where suitable.

Much of the criticism directed at reviews seems to stem from the sometimes muddled approach of the reviewer. It is important for the review writer to decide whether his review is to be a critical appraisal or simply a summary of current advances. He must be quite certain whether he is writing for the specialist in the subject, or for the man working in a related subject, or for the layman. It is probably the ideal to have frequent summary reviews of current advances and less frequent comprehensive critical reviews. The answers to Question 7 indicate that in general the criteria for good reviews are being met, despite the fact that many bad reviews are still published. Only 14 chemists indicated their general dissatisfaction with the present state of review writing. There was also general satisfaction with the coverage of subjects in reviews. Very few chemists gave examples of subjects which they felt were not adequately reviewed. The following were thought to be the main deficiencies: (1) organic methods of analysis, (2) laboratory techniques, (3) solid state chemistry (not physics), (4) application of natural sciences to archeology, (5) fermentation, (6) theoretical spectroscopy. (See Appendix 4, Table 2.)

Failure to review the patent literature was also mentioned as being a common deficiency, and neglect of the history of the subject being reviewed was also instanced. A more serious complaint was failure to survey literature published in the lesser known languages, particularly Russian, Czech, etc.

The answers to Question 10 show an overwhelming majority in favour of confining reviews to review journals and against having journals publishing original work interspersed with review articles. This does not apply to journals which seek to provide a medium for preliminary publications and a general survey of scientific news and progress (e.g., *Nature, Science*). Chemists who preferred the mixed journal pointed out the desirability of having all one's literature in one place, and the advantage of having a review on a particular subject appearing in a journal which reports original work on the same subject. This is a conflict of subject matter versus form but the minority view would be more cogent were scientific journals subject specific. Unfortunately they are not, and one cannot be certain that all work in a specific journal will be on the same subject. A reasonable compromise would be for all journals which publish original work and reviews to confine the reviews to a review section, as many already do.

The meeting together of experts in particular fields to discuss aspects of their work would seem to provide an admirable basis for the publication of reviews of current advances. Today the symposium is an extremely useful means of gathering scientists together for the exchange of information and ideas. It is unfortunate that the published reports of such meetings are not always so useful as one might expect. A number of reports of symposia are too brief to do anything but whet the scientific appetite, and others merely duplicate work which has already been published elsewhere. The most useful reports from the review aspect have the advantage of reporting expert opinion and summarising work in progress in well-defined fields. A major criticism of symposia reports is that they often appear at least a year or more after the meetings have been held and are often published without adequate editing or indexing. The many chemists who find symposia useful express a wish for separate publication rather than for the papers to appear in journals or supplements to journals. The *Discussions of the Faraday Society* were quoted as symposia which were really useful and models of what such publications should be.

As one might expect, the chemist engaged in industry is far more concerned with the availability of the reports of scientific and commercial research organisations than is his academic confrère, and although 43 chemists considered that such reports were useful to their work, only 19 chemists found such reports readily available. A frequent comment was, “I am not sure what is available,” and it is probably true that there is much potentially useful material being issued which is neither widely reported nor abstracted and is therefore not fully utilized. It must be remembered that in the case of industrial reports the question of patents and commercial competition may restrict their dissemination, whilst in the case of reports from government agencies questions of security may act as a barrier to the spread of information. One cannot help feeling that, if something is worth publishing, it should be published through the usual publishing channels, or remain unpublished and reserved entirely for internal use. There is already too great an eagerness to rush into print and to found new journals, and we are reaching a position where every research organisation and every college and institution have their own periodical publications.

In many cases this is quite unjustified on scientific grounds and is done simply for the sake of prestige or so that the institution may have a publication for "exchange" purposes.

Although the cry "no more journals, please" is often heard there is fairly general agreement that a "Bibliography of Chemical Reviews" would be a useful addition to the literature (41 in favour).

It is true that reviews are noted in *Chemical Abstracts,* but they are not separated from abstracts of original work and it would certainly be an improvement if each of the 31 sections of *Chemical Abstracts* had a reviews sub-division. Unfortunately the annual indexes to *Chemical Abstracts* are appearing more than 12 months after the last issue of the volume to which they are the key so that the proposed bibliography which could appear promptly would not be a superfluity. Reviews are often overlooked because there is no simple adequate key to their locations, and the recent venture of the National Library of Medicine (U.S.A.) is an indication of what can be done in this direction. The two issues of *Bibliography of Medical Reviews* have proved immensely useful in the medical field and there is no reason to suppose that a "Bibliography of Chemical Reviews" would not prove equally valuable to the chemist.

A good review contains a good bibliography so that a bibliography of reviews is in large measure a bibliography of bibliographies and as such represents a key to a large volume of literature and is often a quicker means of obtaining a number of references on a specific subject than are the abstract journals.

It is recommended that if such a bibliography is issued, either annually or at more frequent intervals, "chemical reviews" should be given a wide interpretation and the bibliography should include reviews in the field of industrial chemistry.

The answers to Question 15 show little support for the suggestion of more coordination in review publishing with control by learned societies. There is considerable feeling that a learned society would have a stifling influence on review writing, and many feel that there should be an outlet alternative to publishing papers through a learned society. This argument is attractive but duplication and overlapping of reviews undeniably exists, and it is doubtful whether the alleged advantage of complete freedom to publish whatever editors will accept is more important than the wastage inevitably involved in such a system.

Reviews which already appear under the aegis of a learned society should not need control, but more coordination between societies on an international basis would be advantageous. To ensure that all aspects of a subject are reviewed, particularly those which are less popular, and that no subject is neglected, some means of control preferably by editorial consent seems essential. Few people would advocate rigid governmental control, but many of the answers to this question seem to the author of this paper to indicate an addiction to laissez-faire and a faith in the efficacy of the laws of supply and demand almost to the point of naïveté. Typical comments on this question were: "There cannot be too many reviews—someone will read them." "If a need exists for a review, someone will write one."

Many reviews appear which are in effect undigested lists of references and although this type of review has its uses, it is rather wasteful of publishing space. If one could find an alternative method of disseminating information which would relieve the pressure on the scientific printing presses, the lot of the chemist would be a happier one and we might expect scientific research to proceed more smoothly and gather greater momentum. Unless modern methods are brought into use and unless some rationalisation takes place, the whole of our scientific effort is likely to become polarised by its own productivity. The printed page as a method of conveying ideas has a long history; it may well be, that it is no longer the most efficient way of conveying facts to scientists. Specific answers to specific questions can probably be given more efficiently by means of electronic machines, and it was with this possibility in mind that the suggestion of a documentation centre was raised in this questionnaire. If we had documentation centres staffed by specialists and equipped with electronic computing machines and if the university and industrial laboratories were in telecommunication with these centres, then surely one could dispense with a great deal of printed information. Such a development would certainly obviate the need for annotated reference reviews, and it would be a most useful supplement to scientific publication generally.

No answer to Question 16 envisaged such an ambitious scheme, but about half of those chemists who answered this question (i.e., 34) were in favour of a documentation centre. The main criticisms were the expense involved and a doubt whether it could be staffed by persons qualified to select references.

Expense is certainly a major consideration, and unless all scientific publishing including the primary publication area was reorganized and an examination made of what information could usefully be stored electronically at these centres then such a scheme would be uneconomic. Nevertheless automation in the documentation field is steadily increasing, and in ten years time the improvement in electronic devices, a change in the climate of opinion, and the sheer necessity of avoiding chaos will probably ensure the establishment of such centres.

No one doubts the necessity for a specialised staff at such centres, and it is an all too common prejudice held by the research man at the bench that the specialist in documentation is, or necessarily must be, a clerk. The answers to this question show that the information scientist, that is, the scientist who has specialised in library and information work, is comparatively unknown. This is less true in the industrial field, but in the more rarefied air of academic research the library chemist does not have the status, nor is he thought to be so essential as the laboratory chemist. This attitude is one which, if it persists, might very well cause a breakdown in the whole system of information storage and retrieval, for it is essential, if we are to use the most efficient methods of organizing scientific literature in a scientific way, that we encourage first class scientists to specialise in documentation, library, and information techniques. It cannot be over emphasised that those who record, retrieve, and disseminate either information about or the results of scientific discovery are as vital to the community as the discoverers themselves.

APPENDIX 1 THE QUESTIONNAIRE

INTERNATIONAL CONFERENCE ON SCIENTIFIC INFORMATION 1958 REVIEW LITERATURE QUESTIONNAIRE

Q.1. Please indicate a) your general field of research
b) your specific interest in that field

A.1. a)
b)

Q.2. Do you regularly read reviews a) in your own field
b) in related fields
c) of a general nature

A.2. a) b) c)

Q.3. In your field are there being published a) too few reviews
b) too many reviews
c) an adequate number

A.3. [Please put a tick against either a), b), or c).]

Q.4. Do you find the *Advances in*—and the *Annual Reviews of*—series useful? Have you any criticism or comment to offer?

A.4.

Q.5. Name any review journals (including the type mentioned in the previous question) which you read regularly.

A.5.

Q.6. What in your opinion are the essentials of a good review?

A.6.

Q.7. Are your criteria for good reviews being met at present?

A.7.

Q.8. Can you suggest any improvement in review publication?

A.8.

Q.9. Is there any section of the chemical field which, in your view, is not covered adequately by reviews? If so please give details.

A.9.

Q.10. Do you prefer journals to carry original work and review articles or do you prefer reviews to be carried only in review journals?

A.10.

Q.11. Do you find reports of symposia valuable as reviews.

A.11.

Q.12. Do you prefer symposia to be published a) in journals
b) as supplements to journals
c) as separate publications

A.12. [Please tick either a) b) or c).]

Q.13. Do you consider that scientific reports of individual research organisations are useful to your work?

A.13.

Q.13a. Are they readily available to you?

A.13a.

Q.14. Would you favour the publication of a "Bibliography of Chemical Reviews" on an annual basis?

A.14.

Q.15. Would it be advantageous to have any form of coordination of review publication with control exercised by, say, learned societies? [Please comment.]

A.15.

Q.16. Do you consider a central documentation service which could supply selected references in particular fields to be an alternative or a useful supplement to review publication? [Please comment.]

A.16.

FURTHER COMMENTS

You are invited here to make any comments or suggestions on the whole topic of review literature and its publication which you feel would be helpful and useful to the Conference.

APPENDIX 2 COMMENTS FROM REPLIES

QUESTION 4

Do you find the *Advances in*—and the *Annual Reviews of*—series useful?

Serial No.	
27	As author of analytical chemistry reviews in *Annual Reports on the Progress of Chemistry* for the past three years I have found great difficulty in covering the literature adequately in the space permitted for this review. Naturally the review is biased towards my own interests in this field.
41	These publications generally deal with a narrow field. If it is one's own field, the original literature is usually available and more useful. If not the review is not wide enough in scope to be of value.
48	The *Advances in*—type of review could well be extended to many other subjects than those available at present. Unfortunately most of these are published in U.S.A. and are prohibitively expensive.
50	*Annual Review of Physical Chemistry* is too closely akin to a list of abstracts or references. It is not opinionated enough. Each article would be improved by productions in dual form. (A) Giving a classified bibliography of the year and (B) Reviewing chosen fields, perhaps in alternate years or even less frequently and setting the advances of a couple of years in perspective.

QUESTION 8

Can you suggest any improvement in review publication?

Serial No.	
7	Reviews should be split into two classes: *(a)* Exhaustive and detailed reviews; the only type of any use when work is to be commenced in a given field. *(b)* More superficial and readable type of review of the “keeping up” kind. There should be separate journals for *(a)* and *(b)*.
8	Authors should be clear on the nature of the intended readership. Is the review intended to provide a manual for the new research worker or an aid to the expert? Is it intended to attract interest in the topic, or in fact to survey recent advances? Mixing of these aims can lead to a review of low value to each particular group.

Serial No.	
10	The review most useful to me in the general chemical field is that in which specific reactions and compounds are dealt with and which lists all the examples in the literature. This type of review is not encountered so frequently in British publications although it is only fair to say that many articles of this type may be published as books.
11	Quality might be improved if the publishing body solicited contributions from recognised authorities personally. Usually it is left to the authors to make the overture—and the wrong authors do it.
20	Reviews should be referred by two persons, one an expert in the field to check veracity and general scope of the article; the other should not be working in the field and he should be responsible for ensuring that the review is intelligible to a wide public.
21	Some reviewers tend to review their own work in a subject in detail and only outline other contributions. U.S.A. reviewers sometimes neglect U.K. journals.
28	I would suggest a separate journal of reviews in analytical chemistry with more chemical journals accepting review articles as well as original papers.
31	A regular series of reviews on specific topics of a more general character to cover a definite period of time, say last 5 or 10 years.
32	It might be suggested that all reviews should be written by retired scientists in their own field who have maintained touch with their subject. There are too few good reviews and too many bad reviews.
42	There is a requirement for more reviews of greater length and greater authority. Perhaps something approaching a short monograph. The advantage of publishing monographs in this form is that the potential reader or user is more likely to see them and read them if they appear in a standard journal than if they are published as a book.
43	One is tempted to suggest streamlining them all into one publication, but the present hotch-potch gives a variety of presentations and breadth of cover which avoids staleness. It has arisen by the machinations of supply and demand and is probably a good thing.
44	There is too great an assumption by many writers that they need only summarize material already familiar to readers. On the other hand some U.S.A. reviews are too long; sometimes verbose. *Quarterly Reviews* probably sets the highest standard, especially in the editors' aims. It would be valuable if one publication (*Annual*

Report on Progress in Chemistry seems most appropriate) concentrated on summarising work done in a particular year and other journals attempted to give a more connected account of "recent progress." The latter need not be done in terms of particular dates, but according to the needs of the subject. This could still be done in a journal appearing annually.

50 Editors should encourage authors to write reviews in pairs: single-handed efforts make an unfair demand on a man in these days of an exuberant literature.
Authors of comprehensive reviews should be encouraged to give a roughly classified appendix of articles impinging on their subject where inclusion however would make balance difficult to achieve or where merit is too slight to be made much of in the text.
Editors should demand reasonably detailed plans and feel free to request authors to include fields or aspects which might seem neglected or to give reasons for not doing so.

QUESTION 10

Do you prefer journals to carry original work and review articles or do you prefer reviews to be carried only in review journals?

Serial No.

48 Occasional reviews in journals publishing original work are useful. I consider the method of *Industrial and Engineering Chemistry* of issuing a review edition once a year, which contains reviews written by specialists in certain fields, an excellent idea. These are for example more easily read than those published in the *Annual Reports on the Progress of Chemistry*.

QUESTION 13

Do you consider that scientific reports of individual research organisations are useful to your work?

Serial No.

57 It is difficult to obtain decent coverage of many of these reports, the majority of which are not covered by *Chemical Abstracts*. This makes location of information unnecessarily difficult. If a paper is worth publishing it should have wide circulation. There appears to be too great a haste to rush into publication in some quarters!

59 They are useful only in providing more rapid access to recent work which is mostly published through the normal channels in the end. There are probably many such reports of which one does not know.

QUESTION 15

Would it be advantageous to have any form of coordination of review publication with control exercised by say learned societies?

Serial No.	
44	I would doubt the wisdom of arranging anything on a formal basis and would prefer cooperation based on the good sense of the respective editors. Unless there are strong reasons against, I think the freedom of editors to commission reviews and of reviewers to approach journals of their choice should be maintained.

QUESTION 16

Do you consider a central documentation service which could supply selected references in particular fields to be an alternative or a useful supplement to review publication?

Serial No.	
7	If each paper were coded by specialists and the chemist requiring references could call for selected code numbers, this might work. It would require minute coding and several code numbers would apply to most individual papers and the sorting would have to be done by Hollerith or similar machines.
11	The idea of *Beilstein* operating on a postal basis is intriguing and might be useful because the questions it could be asked would be direct and specific.
27	No service could exist on an economic basis to cover the vast literature which we now have. For example to cover the Russian literature alone would be costly enough. American *Chemical Abstracts* are coming against this problem and shortly will price themselves out of the market.
38	I think it might be invaluable. Many librarians at present do supply information and in effect do some library research for me, but they vary in the time and effort one can reasonably ask of them, if any, and the present situation is rather haphazard.
44	The best thing in my own field is Deitz, *Bibliography of Solid Adsorbents,* published by the National Bureau of Standards. This is tremendously valuable as a supplement to reviews and a model of what might be done in other fields. I can hardly rate its value too highly.
51	Selected references cannot be an alternative to a review. One of the prime merits of the review is that it can be read and digested at leisure and that it often stimulates work and reading in related fields.

FURTHER COMMENTS

Serial No.	
6	I should like to see in this country a review journal devoted exclusively to book reviews. It should be produced as cheaply as possible but should publish reviews quickly. Most book reviews are far too slow in appearing and too few foreign books are reviewed.
7	I feel that reviews should be written under a "contract" entered into before writing begins. Initial suggestions for a review could come either from author or editorial board, but both should be in agreement about the approximate character of the review before it is written and unless it is obviously unsuitable there should be virtual certainty that the review will be printed. Otherwise those with many demands on their time simply cannot gamble with the time and effort needed to write reviews.
10	Much time is occupied in studying original literature. Any means of reducing this task is of great value and the publication of reviews is probably the most important facet. The preamble to original articles is a further method (particularly in American journals where space is less limited) but of less value. The only alternative would be a central documentation service (punched card or similar), but even this would not replace the review especially as a means of disseminating scientific knowledge to those people who are searching for original ideas.
11	Resist any attempt to introduce yet more journals. There are too many already. Concentrate on improving the existing ones.
13	Apart from *Annual Reports* review articles probably arise through the author's own feeling of a need for a review in a particular subject. This method may be a bit haphazard but is probably a reasonable one, and may account for a spontaneity and freshness in review articles which might be lost if more rigid control of articles was exercised.
20	It would be helpful if reviews could be really up to date. A good deal of delay is due to the fact that authors try to write reviews in the "spare time" when they are not lecturing, demonstrating, or doing research. The only solution I feel is for the Institution employing the reviewer to grant him a Sabbatical term.
22	I think the question of one worker writing several reviews more or less simultaneously for different publications requires consideration. If a subject is worthy of review in a number of places, then different authors should be chosen.

Serial No.	
24	Undoubtedly the biggest drawback to review literature is the almost inevitable delay in publication which results in the review being somewhat out of date. I feel that this point may be important since it is almost physically impossible to read all the spate of original literature and conduct research at the same time. Solution of this problem is difficult and I cannot pretend to know the answer except for suggesting that more frequent publication of reviews in somewhat narrow fields may help. This, however, would seem to defeat the object of a general review which should be to give a broad picture of advances in a particular field of research.
31	Review literature can serve two purposes—as an aid to teaching and to research. The type of article required for the two purposes would be different. Those for teaching should be of a character to be of use to the advanced undergraduate, much less detailed and of wider background than those applicable to research fields.
32	The main purpose of a review is not to bring one up to date in one's own field; every senior chemist should be on top of his subject. As an introduction to newcomers reviews are useful if authoritative and also for keeping in touch with related fields of work.
	A serious objection to many chemical reviews is that they are often pure "scissors and paste" efforts by, at best, a junior lecturer who is usually quite unable to appraise the value of certain publications, e.g., patents, and who relies too much on abstracts for his information.
	I feel the Conference could do a useful service by improving the standard of reviews mainly by eliminating the bad or unnecessary reviews. With the greatly increased output of scientific literature every effort is needed to keep the vast volume of literature within bounds so that the task of the reader is rendered easier and he is not confused by endless repetition.
35	I find the most useful form of review is one with a short introduction with reference to major contributions, the bulk of the review covering the more recent history systematically and comprehensively indicating in very general terms the material which may be found in the references quoted.
37	A yearly index covering all journals in a specific field would be useful.
41	As a teacher the general type of review is of more value to me than that which deals in detail with some highly specialised topic. I use reviews almost solely as a means of keeping up to date in branches of physical chemistry other than my own. As far as re

Serial No.	
	views in my own field are concerned I find them of more use as an index to recent literature than as expositions of their topics.
42	A greater distinction should be made between reviews which (A) attempt to give complete coverage of the literature in some restricted field, and (B) those which attempt to give a more general account of recent advances in some (perhaps wider) topic. (A) may become essentially a list of references; if so it is useful and interesting only to the expert in that field or to someone wishing to specialize therein. (B) Must be authoritative and must go back sufficiently far for the more general reader to appreciate what the recent advances really are and must show their relation to other topics. Authors must be allowed sufficient length to do this effectively otherwise there is a risk that the article will be trivial to the specialist and meaningless to the non-specialist.
43	The main use of reviews is in my view by people having to obtain quickly an overall picture, as for example when starting new work. For this they are invaluable. Apart from this they enable one to keep abreast of subjects fringing on one's own or having some personal interest. They are not all that useful in the specialist's own subject except to obtain other people's general views and to check that nothing has been missed.
44	I would suggest that more reviews should be aimed at the undergraduate. It might even be worth converting a review journal to this purpose. *Journal of Chemical Education* is valuable but is rather bitty and there is nothing of its kind in this country (Great Britain) above the level of *School Science Review*.
48	Journals of the type *Quarterly Reviews* could well be encouraged and extended even to the extent of taking larger review articles. Such reviews are reasonably priced and easily available to Fellows of the Chemical Society. On the other hand *Advances in*—publications are also excellent and are able to devote much more space. Developments on these lines at a cheaper cost would be useful.
52	A good review requires a very thorough search of the literature as well as critical examination of the relevant papers. This plus the time involved in producing a well-constructed review make review writing very laborious and time-consuming. From this point of view it would seem that too many reviews are attempted. On the other hand "reviews" which are little more than a collection of references with brief abstracts are of value providing they are comprehensive.

Serial No.	
53	A review is a survey of the status of continuing scientific work. It must inevitably contain the reviewer's creative assessment. He must reveal controversy and conflict, but he may state his own opinions. He must if possible give the feeling of living people striving to solve problems in science. It is important that a review should cover all the work in its field and not confine itself to any set headings. Then when a new review is required to cover a new subject, there should not be any difficulty in discerning the need.
56	Position generally satisfactory so far as my own interests are concerned. Where gaps appear it is quite possible to bring these to the notice of editors or publishing committees who can then solicit reviews if they consider this desirable.
60	It seems to one that the "topic" usually adjusts itself according to the laws of supply and demand. The situation is perhaps, not ideal, but I see nothing wrong with it. Other matters are in far more chronic state: e.g., the number of journals there are; their non-specificity, making the fraction of their contents useful to workers in one field too small to make it worth buying them; and the non-speed of publication.
63	The whole topic of review literature can be summed up in this contradiction, viz., that reviews, detailed, clear, comprehensive, frequent, and accessible, are very desirable and yet because of the shortage of time and the frailty of human nature, the more detailed, comprehensive and numerous they become, the more oppressive is the burden they comprise. Our outstanding problem is not the shortage of reviews, even of good reviews, but the shortage of time to read and absorb them.
65	I would commend the *Commonwealth Bureau of Agricultural Sciences* as a model which could perhaps usefully be followed in other branches of chemical research where review and bibliographic service are wanting.

APPENDIX 3 JOURNALS CITED IN ANSWER TO QUESTION 5

1. *Advances in Agronomy*
2. *Advances in Cancer Research*
3. *Advances in Carbohydrate Chemistry*
4. *Advances in Catalysis*
5. *Advances in Enzymology*
6. *Advances in Food Research*
7. *Advances in Protein Chemistry*
8. *Analyst*
9. *Analytical Chemistry (Annual Review* issue)[1]
10. *Angewandte Chemie*
11. *Annual Reports on the Progress of Chemistry*[1]
12. *Annual Review of Biochemistry*[1]
13. *Annual Review of Microbiology*
14. *Annual Review of Nuclear Science*
15. *Annual Review of Physical Chemistry*[1]
16. *Annual Review of Plant Physiology*
17. *Bacteriological Reviews*
18. *Biochemical Society Symposia*
19. *Chemical Engineering Progress*
20. *Chemical Reviews*[1]
21. *Chemistry and Industry*
22. *Commonwealth Bureau of Agriculture Publications*
23. *Discussions of the Faraday Society*
24. *Endeavour*
25. *Fortschritte der chemischen Forschung*
26. *Industrial and Engineering Chemistry* (*Annual Review* issue)
27. *Modern Packaging Encyclopaedia*
28. *Nature*
29. *New Scientist*
30. *Nutrition Abstracts and Reviews*
31. *Organic Reactions*
32. *Physical Review*
33. *Physiological Reviews*
34. *Progress in Nuclear Physics*
35. *Progress in Organic Chemistry*
36. *Progress in Stereochemistry*
37. *Quarterly Reviews*[1]
38. *Reports on the Progress of Applied Chemistry*
39. *Reports on Progress in Physics*
40. *Research* (articles in)
41. *Review of Scientific Instruments*
42. *Review of Textile Progress*
43. *Reviews of Modern Physics*
44. *Reviews of Pure and Applied Chemistry (Australia)*

[1] Journals most frequently mentioned.

45. *Royal Institute of Chemistry, Lectures, Monographs, and Reports*
46. *Schimmel Berichte* and other house journals
47. *Science Progress*
48. *Soap and Chemical Specialties Yearbook*
49. *Times Science Review*
50. *Vacuum* (articles in)

APPENDIX 4 TABULATION OF DATA AND KEY

KEY TO THE TABULATION

Each completed questionnaire has been given a serial number which is used in Appendix 2 as a means of identifying authorship of comment. "See Appendix 2" appearing in a column of the tabulation indicates that a comment was made which can be found under the number of the question and against the relevant serial number in Appendix 2.

All scientists represented by the completed questionnaires have been categorized by the type of Institution in which they are employed according to the following code:

Aac	Agricultural College
Ap	Polytechnic
At	Technical College
Au	University
Auc	University College
Aum	University Medical School
G	Government Laboratory
I	Private Industry
In	Nationalised Industry

In all answers the following code applies:

Y	An affirmative answer
y	A modified affirmative, e.g., "sometimes," "might be useful"
N	A negative answer
—	No answer given

All periodicals are referred to by the running number assigned to them in Appendix 3. This applies to the answers to questions (4) and (5).

QUESTION 6. What in your opinion are the essentials of a good review?

Code 1 Good introduction, defining scope

2 Clear, readable, concise

3 Critical; authoritative

4 Comprehensive coverage

5 Selective coverage; adequate coverage

6 Up to date

7 Bibliography: (7a) full references; (7b) selected references

8 Use of tables, diagrams, etc.

9 Well balanced; good correlation and integration of topics

10 Intelligible to non-specialist

Some answers to this question divide reviews into General and Specialist. In these cases *(a)* indicates the criteria desired for general reviews and *(b)* the criteria desired for specialist reviews.

QUESTION 10. Do you prefer journals to carry original work and review articles or do you prefer reviews to be carried only in review journals?

Code S Indicates a preference for reviews in separate journals

O Indicates a preference for reviews and original papers in the same journals

NP No preference

QUESTION 15. Would it be advantageous to have any form of coordination of review publication with control exercised by say learned societies?

Code 1 It is useful to have differing viewpoints of the same subject

2 It is impracticable

3 It should be left to the editors concerned

4 There is little duplication of reviews

5 There might be prejudice on the part of controlling bodies

6 It would avoid duplication and ensure coverage

QUESTION 16. Do you consider a central documentation service which could supply selected references in particular fields to be an alternative or a useful supplement to review publication?

Code Y Yes; alternative or supplement not specified

Ya Yes, as an alternative

Ys Yes, as a supplement

1 Searching the literature should be a scientist's personal responsibility

2 It would merely duplicate *Chemical Abstracts* or *Current Chemical Papers*

3 It would be impracticable; or too expensive

4 Industrial libraries, etc., already have such a service

5 Non-specialist selection would be unreliable

FURTHER COMMENTS

Comments made in this section appear at the end of Appendix 2 under the heading “Further Comments” and against the serial number of the questionnaire in which the comment appears.

TABLE 1 Tabulation of data, review literature questionnaire, Questions 1 to 8

		1	2			3	4		5	6	7	8
Serial No.	*Category*	a) *General interest* b) *Specific interest*	*Reading of reviews* (a)	(b)	(c)	*No. of Reviews*	*Annual Reviews and Advance series* Code	Comment	*Journals cited*	*Criteria*	*Criteria fulfilled*	*Improvements needed*
1	Au	a) Organic b) Carbohydrates	Y	y	y	c	Y	Extremely valuable	3;11;12	9: Emphasize the novel or unusual	Y	More detail
2	Au	a) Organic b) Carbohydrates	Y	Y	N	c	Y	—	3;11;12; 15;37	2;3;8	Y	N
3	Au	a) Organic b) Carbohydrates	Y	Y	y	c	Y	—	3;11;37	2;5;6;7;9; 10	y	N
4	Auc	a) Organic b) Alkaloids; terpenes	Y	Y	y	c	Y	11, Very useful	10;11; 20;36;37	a) 2;5 b) 4;7	Y	N
5	Auc	a) Organic b) Alkaloids; porphyrins	Y	Y	Y	c	Y	11 & 35, Especially useful	11;35	—	—	—
6	Au	a) Organic b) Quinones—hydrogen bonding—phenylation	Y	Y	y	c	Y	They appear late, perhaps unavoidably	11;20; 37;47	a) 1;6;7b b) 1;6;7a	Y	Too much overlapping requires international coordination
7	Ap	a) Organic b) Heterocyclics; stereochemistry	y	y	N	—	—	Rarely read these series	11;20; 28;37	4	y	See Appendix 2
8	Auc	a) Organic b) Heterocyclics & organometallics	Y	Y	Y	c	Y	—	11;20; 31;35; 37;45	2;3;7b;8	Y	See Appendix 2
9	Au	a) Organic b) Phosphorus compounds	Y	y	y	c	Y	11, Over compressed	37	2;5;9	Y	N
10	I	a) Organic resins b) Phosphorus & boron chem.	y	Y	Y	a	Y	Extremely useful; particularly as subject index to *Chem. Abs.* is delayed	11;20; 37;38	3;4;6;7	Y	See Appendix 2
11	Au	a) Organic reactions b) Peroxides; mechanisms	Y	Y	N	c	Y	11, Particularly useful. Work in Russian & Japanese, etc., is sometimes overlooked	11;20; 25;37	a) 2;9 b) 4;6;7a;8	Y	Speed up publication See Appendix 2
12	Ap	a) Organic chemistry b) Reaction mechanisms	Y	Y	Y	c	Y	Good surveys of work in particular fields	20;37	a) 1 b) 7;9	Y	N
13	Au	a) Organic chemistry b) Reaction mechanisms	y	y	N	c	Y	Extremely useful	11;37	3	—	N
14	I	a) Organic chemistry b) Polymerisation	Y	N	N	c	Y	—	15;37	2;5	Y	N
15	I	a) Plastics b) Monomers	Y	Y	Y	c	Y	—	11;20; 37	1;3;5;7	Y	N

16	Au	a) Organic b) High polymers & dyes	Y	Y	Y	c	Y	—	42	5	Y	N
17	I	a) Organic b) Acetylene reactions	y	y	y	c	Y	—	—	—	—	—
18	Auc	a) Organic b) Aliphatic chemistry	Y	Y	Y	c	Y	11, Sometimes too condensed	11;12; 20;37	2;5;7	y	—
19	Au	a) Organic b) Proteins, peptides	Y	Y	Y	c	Y	Too expensive	7;11; 12;37	—	Y	Sale of separates
20	Au	a) Biological b) Proteins & nucleic acids	Y	Y	Y	c	Y	Some try to cover too much literature	7;11;12; 20;37	2;3;5;7b	y	See Appendix 2
21	I	a) Biochemistry b) Enzyme reactions	Y	Y	Y	a	Y	5, Very useful; *Ann. Revs.* useful only for references	5;6	2;5;9	y	Sale of separates, see Appendix 2
22	Aum	a) Enzymology b) Sulphatases	Y	Y	Y	c	Y	Very useful	5;12; 16;33	7;9	Y	N
23	Aum	a) Cancer research b) Chemical carcinogenesis	Y	Y	y	c	Y	—	2;11; 12;18	1;3;9	N	Too many reviews are just lists of references
24	Au	a) Fermentation b) Cultivation of micro-organisms	Y	Y	y	a	Y	Out of date on publication	5;13;17	5;6;7	Y	More critical presentation needed
25	I	a) Analytical b) Methods	Y	Y	Y	c	—	Rarely used	8;9;37	—	—	—
26	I	a) Analytical b) New techniques	Y	y	y	c	y	—	—	—	—	—
27	Au	a) Analytical b) Classical organic & inorganic	Y	y	N	b	Y	See Appendix 2	9;11;38	3;4;7	y	Credit is not always given to original authors of work cited
28	Au	a) Analytical b) Ion exchange & chromatography	Y	Y	Y	a	Y	Tend to be just a list of references	11;15;37	5;9	y	See Appendix 2
29	At	a) Analytical b) Organic chemistry	Y	Y	Y	a	Y	See Question 8	9;11;15; 20;35;37	4;6;7;9	Y	Annual review of organic chemistry would be useful
30	At	a) Physical-organic b) Polarographic studies	N	N	Y	a	Y	11, Invaluable for teaching	11;23;37	2;4;7a;9	Y	N
31	Auc	a) Organic stereochemistry b) Optical activity	Y	Y	Y	c	—	—	11;20;37	3	—	See Appendix 2
32	I	a) Organic b) Physical-organic & analytical	N	y	y	a	y	*Advances*, out of date; *Ann. Revs.*, often good	46	1;2;3;7;8	N	See Appendix 2
33	Auc	a) Stereochemistry b) Optically active compds	Y	Y	Y	c	Y	I look through all that are available to me and read those bearing on my subject	see previous column	3	Y	Speed up publication
34	In	a) Coal carbonisation b) Industrial waste treatment	Y	Y	Y	c	N	Reviews in ordinary journals are adequate	N	3;4;6	Y	Fuller information about Russian & Czech, etc., papers needed
35	I	a) Organic b) Petroleum	Y	Y	N	c	Y	—	9;11;26; 37;38	2;7;9	Y	I find American reviews best

		1	2			3	4		5	6	7	8
Serial No.	*Category*	a) *General interest* b) *Specific interest*	*Reading of reviews* (*a*)	(*b*)	(*c*)	*No. of reviews*	*Annual Reviews and Advance series* *Code*	*Comment*	*Journals cited*	*Criteria*	*Criteria fulfilled*	*Improvements needed*
36	I	a) Pharmaceutical b) Cosmetics, etc.	Y	Y	N	c	Y	They are well done and considering their scope published promptly	30;37	2;7b Chronolog. treatment of subject	y	Some reviews quote no references; some become a list of names and work done
37	I	a) Cereal chemistry b) Flour milling	Y	Y	Y	c	Y	Not enough attention to patent publications	3;5;7; 26;46	3;5	y	More cumulative indexing; use of "briefs"
38	Aac	a) Soil analysis b) Dairy chemistry	Y	y	y	c	Y	—	11;22;45	1;2;3;8	Y	—
39	I	a) Packaging b) Aerosols	Y	Y	Y	a	N	Packaging is hardly ever dealt with	27;48	6	y	There are no review journals on packaging in U.K.
40	I	a) Tech. Physical chemistry b) Water treatment	Y	y	y	a	N	—	3;4;9	—	—	N
41	At	a) Electrochemistry b) Metal deposition	Y	Y	Y	c	N	See Appendix 2	11;37;38	2;7	Y	N
42	Au	a) Physical b) Electrochemistry	Y	Y	Y	a	Y	Some tend to degenerate into lists of references	11;15; 37;39	see Appendix 2	N	See Appendix 2
43	I	a) Precipitation of hydroxides b) Albuminas	Y	Y	y	c	Y	Useful as aides-memoirs for superficial coverage	9;19; 26;29; 37;49	2;4;7b	y	See Appendix 2
44	Au	a) Physical & inorganic b) Surface chemistry	Y	Y	Y	c	Y	*Advances*; very specialised; *Ann. Revs.*, neglect subjects which are not advancing rapidly	4;11; 15;20; 26;37	1;3;7 Exposition of new concepts	y	See Appendix 2
45	Au	a) Physical b) Surface chemistry of solids	Y	Y	y	c		*Advances*, best *Ann. Revs.*, tend to be disjointed	37	2;3	N	Improvements in readability needed
46	Auc	a) Solid state (1), chemisorption (2), physical adsorption (3)	Y	y	N	a(3) b(1) c(2)	Y	*Ann. Revs.*, too broad and too frequent	4;20; 32;43	3;7	Y	Greater coordination between review journals; greater planning and indication of future proposals
47	Ap	a) Physical b) Refractivity & surface tension	Y	Y	Y	c	Y	—	20;37;47	2;7	Y	More frequent publication
48	Au	a) Physical b) Colloids	Y	Y	Y	a	Y	See Appendix 2	7;9;11;26; 37;40;50	a)2;7;10 b)2;6;7	Y	Speed up publication; cheaper cost
49	Auc	a) Electrochemistry b) Non-aqueous solutions	Y	Y	y	c	Y	—	11;15; 20;37	a)2;7b b)3;4;7a	Y	Chemistry well served

50	Au	a) Physical b) Flame & explosions Free radical reactions	Y	y	Y	c	y	See Appendix 2	11;20;37	1;2;3;9;10	Y	See Appendix 2
51	Ap	a) Physical b) Kinetics	—	Y	Y	a	y	Tend to become a catalogue of papers	20;37	10; it should indicate desirable developments	N	See Appendix 2
52	Au	a) Kinetics & spectroscopy b) Free radicals, & gas reactions	Y	Y	Y	c	y	Useful as a source of references; tend to be a collection of uncorrelated summaries of papers	11;15; 20;37	3;9	N	More reviews on topics & techniques which are developing
53	At	a) Physical b) Heterogeneous catalysis	Y	Y	Y	c		Prefer *Advances; Ann. Revs.* too cursive	4;11;15; 20;37;41	2;3;9	Y	A need for large views in brief,e.g., a review of reviews
54	At	a) Physical chemistry b1) Surface chemistry b2) Boron, organic compds.	Y	Y	Y	b1)a b2)c	Y	More practical details needed	4;11;20; 37;47	3;9	y	More journals could be published
55	Auc	a) Inorganic b) Boron & organometallics	Y	Y	Y	c	y	Useful as bibliographies not as reviews	11;20;37	3;6	Y	N
56	Au	a) Inorganic; structural b) Thermochemistry & solid state	Y	Y	N	c or b	Y	—	11;15; 20;37; 44	a)5;10 b)4	Y	Some popular topics appear too often
57	G	a) Inorganic b) Purification & analysis	Y	Y	Y	c	Y	Useful in so far as one's own field is related to other topics	9;20;37;44	3;5;9	Y	—
58	Au	a) Physical & radiochemistry b) Kinetics, photochemistry	Y	Y	Y	c	Y	There is overlapping which perhaps is a good thing	4;11;14; 15;20; 37;39	a)1;5;6 b)4; need not be critical	y	N
59	Au	a) Radiochemistry b) Measurements & tracer studies	Y	Y	y	c	Y	Too many degenerate into annotated bibliographies	11;14;15; 34;37	3;4	N	More critical assessments needed
60	Au	a) Physical b) Spectroscopy	Y	y	y	c	Y	Very useful	11;15;37	2;3;5;9	y	N
61	Auc	a) Physical b) Molecular spectroscopy	Y	y	Y	c	Y	By their nature they consist of little more than strings of references woven about a central theme	—	3	Y	Speed up publication
62	Au	a) Chemical spectroscopy b) Infrared & nuclear magnetic resonance spectra	Y	Y	y	c	Y	15, Too detailed for reading all through	15;20;37	a)2;5 b)4	Y	N
63	Au	a) Structural b) X-ray crystallography	Y	Y	Y	c	Y	—	11;37	1;2;9	Y	Space restrictions often cramp style
64	Au	a) Theoretical b) Quantal & statistical	y	N	Y	c	Y	—	37;43	3;6;9	Y	Need for more critical discussions of topics
65	Au	a) Soil & plant b) Mineral nutrition	Y	Y	Y	c	Y	Valuable to teachers also	1;11; 21;24	3;7	Y	N

TABLE 2 Tabulation of data, review literature questionnaire, Questions 9 to 16

	9	10	11	12	13	13a	14	15		16	
Serial No.	*Inadequate review coverage*	*Separate Journals*	*Symposia*		*Scientific Reports*		*Annual Bibliography of Reviews*	*Coordination and control*		*Documentation centre*	
								Code	*Comment*	*Code*	*Comment*
1	N	S	Y	c	—	N	Y	N	1	N	1
2	N	S	Y	c	N	N	Y	N	2;3	N	2
3	N	S	y	c	y	y	N	N	2; only of value if international	N	—
4	N	S	Y	c	N	Y	N	N	4	N	2;3
5	N	S	y	b	N	y	N	Y	Reviews tend to proliferate unduly	Ys	It would save many weary hours of literature searching
6	—	S	y	a	Y	N	N	N	Overlapping is mainly with foreign reviews	N	1;4
7	Macromolecules	S	N	b	N	N	N	Y	Some sort of rationalisation is necessary	y	See Appendix 2
8	Heterocyclics other than natural products	S	N	c	N	N	Y	N	5	Y	Desirable but 3
9	N	S	N	c	N	N	Y	Y	See Appendix 2	N	1
10	Polymers	S	Y	a	N	N	Y	N	3;4	Ys	It would have to be on the scale proposed by, e.g., Dyson
11	Practical organic chemistry	S	Y	c	N	N	N	N	4	Y	See Appendix 2
12	N	S	N	c	—	y	Y	—	Already exists to some extent	ys	Good reviews with access to literature is sufficient
13	N	S	Y	c	N	—	N	N	3	Ys	Probably 3
14	N	S	N	c	N	Y	Y	Y	6	N	5
15	N	S	y	b	Y	Y	Y	Y	—	N	Unless the centre could supply critical expositions
16	—	NP	Y	a	Y	N	N	N	Doubtful	Ya	—
17	—	O	Y	b	Y	Y	N	—	—	—	—
18	—	S	Y	c	y	N	N	N	Doubtful	N	1
19	—	S	y	c	N	—	Y	N	Doubtful	N	3
20	N	S	N	c	Y	N	N	N	2	Ys	But 3 & 4
21	N	S&O	N	b	Y	Y	Y	Y	Exercised carefully	Ys	—
22	—	S	Y	c	N	Y	Y	N	The more reviews of a subject the better	Ys	5
23	—	S&O	Y	c	Y	N	Y	N	Would lead to less critical reviewing	N	1;5
24	Fermentation	S	Y	b or c	Y	N	Y	Y	6	Y	Would help overcome delays in review publication
25	—	O	Y	a	y	Y	N	—	—	Ys	—
26	N	S	Y	c	Y	Y	Y	—	Already exists for learned society publications	Ys	—
27	N	S	Y	c	Y	N	Y	Y	—	N	3; see Appendix 2
28	Analytical chemistry	O	N	c	y	N	Y	Y	Probably advantageous	N	2

	9	10	11	12	13	13a	14	14	15	16	
Serial No.	*Inadequate review coverage*	*Separate Journals*	*Symposia*		*Scientific Reports*		*Annual Bibliography of Reviews*	*Coordination and control*		*Documentation centre*	
								Code	*Comment*	*Code*	*Comment*
29	Analytical chemistry	NP	y	c	Y	N	Y	N	Learned societies already exercise control on their review publications	Ys	2
30	N	S	Y	b	N	Y	Y	Y	By cooperative action	Ys	—
31	General inorganic chemistry	S	Y	c	—	N	Y	N	Might become more specialised	Ys	—
32	N	S	Y	c	y	N	Y	—	The A.C.S. does well but coordination is more than a part-time job	N	1
33	—	—	Y	c	—	—	Y	N	—	N	5
34	N	O	y	a	Y	Y	Y	N	—	N	—
35	Laboratory technique and apparatus	S	Y	b	Y	N	N	N	Would delay publication	N	4
36	N	O	Y	a	y	y	Y	y	6; avoid stranglehold on authors	Ys	—
37	N	O	Y	a or c	Y	Y	Y	—	More up-to-date industrial matter is needed	Ys	If free from red tape
38	Application of Natural Sciences to Archeology	O or S	y	a or c	Y	Y	Y	y	They should sponsor reviews	Y	See Appendix 2
39	Packaging	O	Y	c	Y	Y	Y	Y	Re-form the Bureau of Abstracts and include Industrial Associations	y	but 3
40	—	NP	Y	b	N	Y	Y	—	—	N	2
41	—	NP	Y	c	N	N	N	N	4	Ys	—
42	N	S	N	b	Y	y	Y	N	No significant advantage	N	2;5
43	Patent literature	S	y	a	Y	N	N	N	4	N	4
44	Colloid science	S	N	c	Y	y	Y	N	3; see Appendix 2	Ys	See Appendix 2
45	—	S	Y	b	y	y	Y	N	—	N	—
46	Solid state *chemistry*	S	y	c	Y	N	Y	Y	See Question 8	Ys	Time saving in literature survey would be great—reviews cannot hope to be comprehensive over wide fields often required
47	—	S	Y	c	Y	Y	Y	—	—	N	—
48	—	O; see Appendix 2	Y	c	Y	y	y	Y	Already exists to some extent	Ys	3; prefer coherent review with suitable bibliography
49	N	S	N	c	N	N	Y	y	For border line fields	ys	But 3 & 4
50	History of the subject in all its aspects	NP	N	b or c	y	N	N	N	3; reviews can be commissioned if required	Y	But it might involve an extravagant misuse of able people

	9	10	11	12	13	13a	14	15		16	
Serial No.	*Inadequate review coverage*	*Separate Journals*	*Symposia*		*Scientific Reports*		*Annual Bibliography of Reviews*	*Coordination and control*		*Documentation centre*	
								Code	*Comment*	*Code*	*Comment*
51	—	S	Y	c	Y	N	—	N	5	N	See Appendix 2
52	Combustion Theoretical spectroscopy	S	N	b	y	N	N	N	3	Ys	
53	—	S	Y	a or b	Y	y	Y	Y	6; see Appendix 2	Ys	A review is a creative dissertation; a supply of references as well would be sheer perfection
54	N	S	Y	c	Y	Y	Y	Y	Coordination between different sciences also necessary, e.g., physics & chemistry	Y	Would have to include borderline subjects
55	N	S	Y	b or c	Y	N	N	N	—	N	1;3
56	N	S	N	c or N	y	Y	Y	N	3; world-wide control impracticable	N	1
57	—	S	y	NP	See Appendix 2	N	N	Y	6	N	—
58	N	S	Y	b	Y	N	Y	y	6; but overlapping not necessarily objectionable	Ys	Not to displace reviews
59	N	S	Y	b	See Appendix 2	y	Y	N	—	N	3
60	N	O & S	Y	b or c	y	N	y	N	3; publishers should give advance information	N	1;2
61	N	NP	y	b	y	N	Y	—	—	Ys	One may learn about work one has missed in one's own field.
62	Theoretical spectroscopy	S	Y	c	y	Y	Y	N	2	Ys	—
63	N	O	Y	b	Y	y	Y	N	Too much coordination can be stifling	Ys	—
64	N	S	N	b	N	y	N	y	But 3	N	3;5
65	—	S	Y	c	y	Y	N	N	Commonwealth Bureau Agr. Sci. is an effective coordinating machine in my field	N	The existing *Bibliography of Soil Science* is adequate

The Place of Analytical and Critical Reviews in Any Growing Biological Science and the Service They May Render to Research

ISABELLA LEITCH

> On my return home, it occurred to me, in 1837, that something might perhaps be made out of this question by patiently accumulating and reflecting on all sorts of facts which could possibly have any bearing on it. After five years' work I allowed myself to speculate on the subject,….—CHARLES DARWIN in the Introduction to *The Origin of Species,* 1859.

A review of literature has long been a conventional part of original scientific papers. Such a review is usually of immediately antecedent work. It is most often narrowly selective and descriptive only of results, conclusions, or opinions, chosen as closely related to the original work to be presented. It need not concern us further.

Reviews of literature, published as such, are now an established part of the information services offered to help scientific research and the application of its results in practice. The stated policy of the Commonwealth Agricultural Bureaux, planned as a comprehensive information service for agriculture, requires the production of reviews. Learned societies and associations of scientific workers produce periodical reviews in their own subjects. The type and quality of such reviews vary widely. Most of them are of the first or second type described below.

TYPES OF REVIEW CLASSIFIED AND BRIEFLY DESCRIBED

THE PERIODICAL REVIEW

Numerically, reviews of this type far exceed any other. Each is prepared by an expert and is concerned with one limited field of research and a limited

ISABELLA LEITCH Commonwealth Bureau of Animal Nutrition, The Rowett Research Institute, Bucksburn, Aberdeenshire, Scotland.

time, most often the year just past. All work published in that time on that subject is to be included. The text consists of brief notes of findings or opinions, with little criticism and usually without analysis or synthesis. Many such reviews are no more than annotated bibliographies.

Even so, they save the time of the research worker who wishes to know what is going on in his own or other fields, and the text is usually sufficient to enable him to select what he wishes to read in the original.

THE OCCASIONAL REVIEW

This review too is usually prepared by an invited expert, but he is not expected to “cover the literature” of any defined period. He is given the opportunity to discuss a subject broadly and to comment on it. (See Information for Authors in *Physiological Reviews*.) He may be selective and critical. In these circumstances the value of an article will depend on the expertise of the writer and the validity of the point of view from which it is written. In practice value tends to vary inversely with the extent to which an article is a summary of opinions and directly with the extent to which it is a discussion of findings.

Physiological Reviews has produced many memorable reviews in this class, of which the following are particular examples which come to mind because they belong to our particular field of work: Henderson (1925) on acid-base balance, Adolph (1933) on water metabolism, Madden and Whipple (1940) on plasma proteins, Soskin (1941) on blood sugar, Kleiber (1947) on body size and metabolic rate, Granick (1951) on ferritin, and Manery (1954) on water and electrolyte metabolism.

THE ANALYTICAL AND CONSTRUCTIVE OR RESEARCH REVIEW

It is concerned directly with facts and findings, seldom with the opinions of the authors of papers from which data are taken. The facts may be of widely different kinds. They may be body measurements, the quantitative results of metabolism experiments, the results of analysis of body fluids or tissues or of food, records of birth and mortality, or, in fact, any biological observations that furnish numerical data which may be treated statistically. Again, the facts examined may themselves be concepts: the concept of being well grown or well nourished, that of optimum requirement for any of the separate constituents of food, of health and normality, all of which will be found to depend on the interpretation of measurements such as are enumerated above. The review of concepts is indeed a more advanced stage of the review of numbers. The two are never completely divorced, for the review of numbers may, and usually does, lead to some new interpretation or concept, and the review of concepts may have as its purpose to clarify issues and show where further numerical information is most urgently required. The rest of this paper will be concerned with the nature of this type of review and its uses, and illustrative examples will be taken from our own publications.

The quotation from the Introduction to *The Origin of Species* which appears on the first page of this paper might have been written of the analytical and constructive review, for which it is difficult to find a short and adequately descriptive name. For brevity, and want of anything better, it may be called the research review, for it is both a review of research and an inquiry into the deductions that may be drawn from an accumulation of results treated as a new whole.

The current short and mostly utilitarian examples have many noble fore-runners in report and book form. It is not possible to take *The Origin of Species* (1859) as the archetype because it was preceded by just over a century by Lind's treatise on scurvy (1753), of which the title page says: *“A Treatise of the Scurvy*. In three parts. Containing an inquiry into the Nature, Causes, and Cure, of that Disease. Together with a Critical and Chronological View of what has been published on the subject.” Of about the same time as *The Origin of Species* there are many of Sir John Simon's reports on the public health when he was chief medical officer to the Privy Council; later came his *English Sanitary Institutions* (1897), Greenwood's *Epidemics and Crowd Diseases* (1935), and that learned and charming book *Ourselves Unborn* by Corner (1944) which offers the unexpected conclusion that “the human body is not notably endowed with specialised anatomical features of a kind that would fit us to perform limited activities supremely well, but on the contrary is built rather closely to the general mammalian pattern, and therefore can perform varied tasks under the guidance of a superior brain.”

THE RESEARCH REVIEW

THE SUBJECT MATTER

The field of biology in general and of nutrition in particular offers plenty of material for treatment in analytical and constructive reviews. The technique can be used for any quantitative problem as soon as enough bits of information are available to provide a worth-while array of data. It may or may not require, or be suitable for, complex statistical treatment. There are two main points to be noticed: the difficulties of assembly because of the scatter of publications, and the complexity of the subject matter itself which, by its nature, accounts for a large part of the scatter and, at the same time, makes assembly and analysis of the composite a necessity.

Publication

A quick look at the number and sort of publications which may have to be consulted will show why assembly is usually a slow and laborious process. When *Nutrition Abstracts and Reviews* was first published in 1931 it contained in the first year titles, or titles and abstracts, of just over 3000 papers from about 400 journals and in about 20 languages. The last completed volume, No. 27, had over 6000 from nearly 600 journals in slightly more languages, and notes on many more papers in symposia or conference proceedings. These 600 journals are seen regularly in five different libraries in seven different places: the Reid Library at the Rowett Research Institute, which is devoted to the literature of animal nutrition, Aberdeen University Library with three branches, the library of the Ministry of Agriculture, Fisheries and Food, the library of the Royal Society of Medicine, and the library of the Lister Institute of Preventive Medicine, the last three in London. In addition, there are reprints from yet other journals and reports from research centres all over the world which are sent direct to the Commonwealth Bureau of Animal Nutrition or the Reid Library, and are not widely available in libraries.

To make an assembly for analytical review, all the available papers must be seen in the original, or photostat copy or other substitute for the original, so that the data if suitable for the purpose in mind may be transcribed. For instance, the 8600 individual observations used in the review of basal metabolism, which will be described in more detail below, were transcribed from roughly 100 papers in medical, physiological, chemical, and agricultural publications issued round the world from Japan, China, Malaya, and India via Europe, Africa, and North America to South America, Hawaii, Australia, and Indonesia. Three diligent young women devoting part of each working day to the task took six months to find, select, and transcribe the individual data. Unless such a team can be mustered, with guidance and linguistic help at need, such a task is beyond the likely powers of individual enterprise.

Complexity

A very large number of variables must be considered and either eliminated by matching or allowed for in the plan of any piece of research in nutrition. First the animals, including man: to be taken into account are breed, sex, age, size at birth and at completed growth, rate of growth or of production (milk, egg, meat), stage of reproductive life, previous dietary history, physical environment (e.g., temperature and humidity). Second, in diet there is a great complexity of nutrients. Known to be of importance are about 20 amino acids, 14 inorganic elements each in many different compounds, and 17 vitamins of known structure; of carbohydrates there are 10 sugars and 8 groups of polysaccharides; 14 fatty acids, and as many triglycerides, innumerable pigments and aromatic substances, as well as many injurious and inert components, present in countless different foods. Add to which that a complete longitudinal study of human growth must take about 20 years, of growth in a pig at least 3 years and even in a rat 9 months; and that there is the reproductive span also, and old age beyond that.

It is obvious that the number of possible questions is enormous and that no one man or one team in the whole of a lifetime can hope to cover by original research more than a minute part of any nutritional field.

Further, nutrition is not so much a science or an art in its own right as a meeting place of many sciences and arts. Since the advent of isotopes in metabolic work, the shepherd and the medical specialist make direct contact with the atomic physicist; the biochemist and the psychiatrist hobnob over esoteric studies of the composition and metabolism of brain cells, and the cook and taste panel consult with big business and public health experts on “additives” to food.

Hence, individual contributions to knowledge in nutrition, and in biology in general, are small and must continue to be so. The contributors are, and will be, scattered over many professions and places of work and their polyglot publications will be similarly spread. It is therefore not only useful, but necessary, that the fragments in a given subject should be brought together from time to time and amalgamated, if money and effort as well as time are to be used in the best possible way. In other words, the making of analytical and constructive reviews seems to us to be essential in biological research, and a very present aid in time of need when briefing is required in medicine, in farming, or in political and economic planning.

CONDITIONS UNDER WHICH ASSEMBLY AND ANALYSIS ARE PROFITABLE

The stimulus to produce a review of this sort comes sometimes from an urgent demand for guidance. For example, the economic depression of the 1930's in Britain, with its poverty and malnutrition while agriculture had surplusses of food it could not sell, called for information on diet in relation to health and the requirements of food and nutrients to maintain health. The prospect of food shortage in 1940 made it compulsory that agriculture in the United Kingdom should use land and stock and feedingstuffs to the best possible advantage, and gave rise to an estimate of the relative efficiencies of farm animals in the production of food for man (see under “Practical Problems, Substitution”). Sometimes a review is called forth by the need to examine an apparently false conclusion in some published work, as, for example, our study of secular change in the height of British adults, and sometimes by mere curiosity, as our analysis of the relation between the weight of mothers and their young at birth. Other instances appear below where we discuss individual reviews.

Once the demand for a review has become apparent, or the necessary stimulus to produce it has been applied, the first question asked is whether enough information exists to provide a worth-while result. The answer to that question one either knows from handling the literature, or must discover by searching.

WASTAGE OF MATERIAL

In the course of assembling data, a certain amount of wastage will always be found, and that for two main reasons.

Faults of Technique

Where the data derive from *ad hoc* physiological or chemical investigation, not much has to be discarded on account of poor technique. Differences between methods are so often small in comparison with the variation between individuals in experiments on animals, or in comparison with what are physiologically significant amounts of the substances being estimated. Where clinical work is concerned it may be necessary to use data from routine clinical examinations, not from research investigations, and then selection may have to be more rigorous. The Medical Research Council Committee on Haemoglobin Surveys (1945) wrote:

There can be little doubt, therefore, that the errors detected in this experiment are smaller than those likely to occur when a relatively inexperienced technician, after forcing a drop of blood from a reluctant patient and using an unstandardised pipette, hurriedly matches the resultant (possibly undergassed) solution in an uncalibrated diluting tube with a colour standard of uncertain value.

That is, no doubt, an extreme indictment, but the investigators who waste time and money making diet surveys by questionnaire or "recall" methods deserve no less severe comment. "No conclusions can be more valid than the data upon which they are based" (Bean, 1948). Similar difficulties arise, of course, in the use of practical feeding experiments, as opposed to planned *ad hoc* research, in animal husbandry.

Faults of Presentation

In these fields, technical, clinical, and agricultural, some of the work and unfortunately an increasing proportion of it, is lost for purposes of resynthesis, because of the mode of presentation. There is, for instance, the paper in which the results of a diet survey are presented, not in absolute terms but as percentages of attainment of some arbitrarily chosen standard. Nothing further can be done with such data.

More important is the overcondensation of reports. Gone are the days when the individual results in an experiment were reported in detail, as in the reports of The Carnegie Institution of Washington and many reports to learned societies. The entry of statistical experts into the biological world, to be welcomed in other ways, has had this effect that results may be given only as a few equations, coefficients, and measures of probability. That is possible because the statisticians have already made sure that the experiments are properly planned to give a statistically satisfactory answer to a defined question. However satisfactory that may be for the solution of the problem set, it is not wholly satisfactory because the work is potentially a small part of a much larger whole. Devices exist by which such results may be combined with similar and similarly planned work, but the data cannot be rebuilt piecemeal into a larger array to serve a different analysis. They are capable of the one interpretation only.

That being so, it is much to be desired that the individual results of biologically important experiments should be tabulated and made available to those who could make further use of them, perhaps through a scheme like that of Dr. Ralph Shaw (1955) for limited editions of highly specialised papers described to the First International Congress on Documentation of Applied Chemistry.

There is wastage also from failure to report all the simple details which must be known or ascertainable in any experiment. For instance, in our assembly of papers on basal metabolism a number could not be used because basal metabolic rate was stated only in relation to surface area, without information about either height or weight; or because age was not stated. The same sort of thing is common in experiments on animals; indeed, in our experience failure to report is the commonest cause of wastage. We suggest that, if experimenters were made familiar with the idea that their results would, in due course, become part of a larger array of data, they would possibly take more thought to describe their material fully.

TYPES OF RESEARCH REVIEW

At opposite extremes in this class are the review which is a purely statistical analysis of assembled data, and that which is a review of concepts, not of primary data.

THE STATISTICAL REVIEW

Basal Metabolism

One example is our study of Basal Metabolism related to Sex, Stature, Age, Climate, and Race (Quenouille *et al.,* 1951). It arose from the request of the Director of the Nutrition Division of FAO to his Standing Advisory Committee at its meeting in Geneva in 1947 for advice on the formulation of energy requirements of different populations for use in relation to estimates of food supplies, actual and desirable. In particular, information was wanted on possible differences with race, size, climate, and the age constitution of populations. There was no review from which guidance on all points could be got, and little information on energy expenditures other than basal that was likely to bear directly on the several questions asked. It was decided therefore to collect information on basal metabolism. A team of three young university graduates searched out and examined papers and, when all the descriptive details required were given, transcribed the data for each individual subject to a prepared form.

Data about climate were provided by Dr. W.B.Fisher, then lecturer in Geography at Aberdeen University, who had been a meteorological officer with the Royal Air Force during World War II, and who still had access to the records of the Air Ministry. The statisticians who made the analysis were Mr. Maurice Quenouille, then lecturer in statistics at Aberdeen University and Mr. A.W.Boyne, statistician to the Rowett Research Institute. They were given no “opinions” about racial or other differences. The inferences are purely statistical.

The points of general biological interest which emerged are: the basal metabolism of a woman is almost exactly seven-eighths of that of a man of the same height and weight; two main racial groups were distinguished, each with two sub-groups; climate produces its maximum effect in hot dry and cold wet areas. Basal metabolism falls at the rate of 3 per cent per decade from age 3 to ages over 80.

It was planned to follow this review with one on the cost of measured amounts of work on the ergometer, of exercise, and then of occupational work, but the results were found to be too scanty and too variable to be of much use. We plan to publish in the autumn a short note, chiefly on the cost of walking.

None of the conclusions about basal metabolism is entirely new, but each is based on a sample so much larger than any used before that it is not surprising that they differ in detail from earlier conclusions, and are more precise. The rate of decline with age is midway between earlier estimates of 2 per cent and

4 per cent derived from small samples. The difference between men and women of the same height and weight is 12.5 per cent; an earlier estimate from a small sample was of a difference of the order of from 5 to 9 per cent. The systematic relation of basal metabolic rate to the climate in which the subject lived is new.

Stature and Growth

Reference has been made above to the apparently false interpretation as immediate stimulus to produce a review. Our study (Boyne and Leitch, 1954) of secular change in the height of British adults was undertaken because Morant's (1950) conclusions on that subject appeared to us untenable chiefly because he treated adult heights as if they derived from a homogeneous population and not from a cohort of survivors over a century in which major changes had taken place in standards of living, public health, and mortality rates. Our reanalysis showed that the British data, on account of limitations of sampling and imperfections of reporting, were in fact inadequate to decide whether a net increase in mean height at completion of growth had taken place or not. It was possible to give clear evidence of an increase of mean height of young adults in Denmark from records of measurements of recruits taken from official statistics and supplied to us by the Director of the Institute of Human Genetics in Copenhagen.

The review of data on adult stature was followed by one on secular change in the height and weight of English elementary school children (Boyne, Aitken, and Leitch, 1957). The data were supplied by School Medical Officers in response to a direct request from us, and a sample of measurements of 1,180,000 children aged 5, 8, and 12 years was built up. Not only did this review put on record for wide circulation important measurements that would otherwise have remained hidden in local medical reports with an exceedingly limited circulation, but, because of the size of the combined sample, deductions could be drawn that would have been impossible from inspection of the individual reports.

Birth Weight

A review of a different type and from a more general biological sphere is of the relation of birth weight of the mother to that of her young (Hytten and Leitch, 1957). The stimulus to produce it came from the suggestion of Dr. R.M.Laws (1956) that the large aquatic mammals, seals in particular, are able to carry relatively heavy young because of the anti-gravity effect of water. He based his obvious misconception on the table on p. 117 of D'Arcy Thompson's (1942) book *On Growth and Form*. D'Arcy Thompson was concerned not with the weight of the young carried, but with the weight at birth in relation to weight to be attained at completed growth. His figures for young born in multiple birth, for instance the bear and the lion, refer therefore to the single members of a litter. Since neither D'Arcy Thompson nor Laws had produced a true picture of the relation of birth weight of young to weight of mother, we decided to look for information about different species. After a search of the immediately accessible journals and monographs we retrieved data for 70 species, and the information was condensed into a graph (Fig. 1).

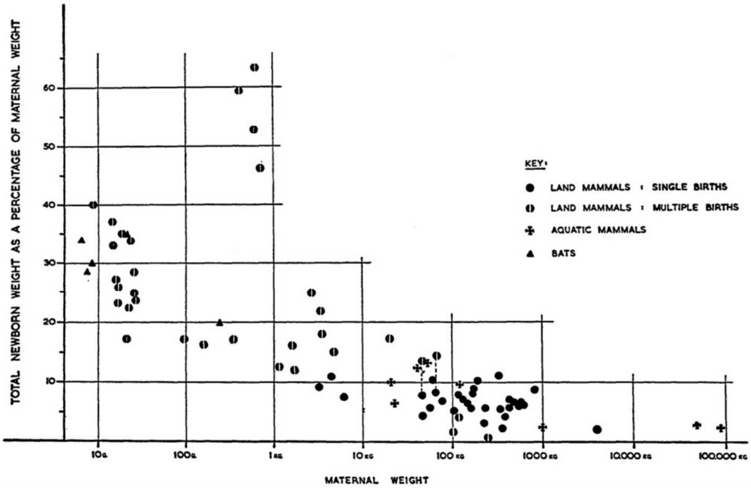


FIGURE 1. The relationship between maternal weight and total weight of the newborn young in 70 mammalian species.

With reference to the aquatic mammals, in the ratio of weight of young to weight of mother they do not differ from large land mammals, including man. The main differences in ratio between animals of the same or similar adult weight are attributable to differences in maturity at birth and to the number of young born at a birth, two or more weighing more than one, but less than would be in proportion to number. This law appears to be valid in general from the bat that weighs 6 g. to the whale that weighs 100,000 kg.

Guineapigs with litters are outside the general picture. We suggest that the domestic guineapig, possibly a descendant of the South American Restless Cavy *(Cavia porcellus),* which produces one, or at most two, at a birth, occupies its strange position as the result of selection for superovulation, and because litters of four have been obtained without much, or any, sacrifice of individual size or maturity at birth.

THE REVIEW OF CONCEPTS

An example of this type of review is that of Thomson and Duncan (1954) on the diagnosis of malnutrition in man. It surveys concepts and definitions, the accepted or proposed criteria of specific deficiency states, the clinical pictures of the well fed and the ill fed, and finally suggests the overriding importance of somatometric studies in which attained dimensions of the body are compared with primary patterns of growth and development. The idea of an inherent pattern in human growth and its importance for health is of recent development (cf. Leitch, 1951), but Hammond and his disciples McMeekan and Pallsson had earlier shown the importance in pigs and sheep of the full development of the “growth potential.”

Other examples from our own publications in which the discussion is chiefly of concepts are the reviews on The Evolution of Dietary Standards (Leitch, 1942), The Calorie Requirement of Adult Man (Keys, 1949), Energy Feeding Standards for Dairy Cattle (Blaxter, 1950), and Iron-Deficiency Anaemia in the Pregnant Woman and Its Relation to Normal Physiological Changes (Hytten and Duncan, 1956).

THE SERVICE OR INTERPRETIVE REVIEW

An early and most distinguished review in this group came from the Rowett Research Institute before there was a Bureau of Animal Nutrition, namely Orr's book on *Minerals in Pasture* (1929). It was the first review to demonstrate the importance of pasture as a crop and the disastrous effect on the health of the grazing animal of lack in the pasture of certain inorganic elements. It was the stimulus to a world-wide programme of analysis and experiment on deficiency of trace elements, which still continues. The review has twice been brought up to date, in 1944 (Russell) and in 1956 (Russell and Duncan).

In this group may be included most of the reviews, other than those listed as reviews of concepts, prepared in or for the Bureau and published in *Nutrition Abstracts and Reviews* or as *Technical Communications*. They include all our earlier reviews of requirements. One on the riboflavin requirement of man by Finn Bro-Rasmussen (1958) of the State Vitamin Laboratory, Copenhagen, has the true pattern of assembly, analysis and reinterpretation. Table 1 of Part II of Bro-Rasmussen's review is reproduced here to show how, by tabulating data from a number of sources, he has been able to establish beyond reasonable doubt that the riboflavin requirement of species ranging in size from the horse to the mouse, and of a microorganism, may be expressed as a function of the energy turnover, but not, as was previously believed, of the protein intake.

There are also collections of data on vitamins in food and one on the water

TABLE 1[a] *Riboflavin requirement of different species related to optimum protein intake and energy intake.*

Species	*Riboflavin requirement* µg. per 100 g. food[b]	µg. per 100 g. protein[b]	µg. per 1000 cal.	*Authority*[c]
Horse	about 350	about 3500	about 850	Pearson, Sheybani, and Schmidt (1944 *a*, *b*)
Man (adult)	—	2000–2500	500–600	Average from values in Table 4.
Calf	about 300	1000–1300	600–800	Brisson and Sutton (1951)
Pig	275–300	900–1000	650–750	Krider, Terrill, and Van Poucke (1949); Miller, Ellis, Stevenson, and Davey (1953); Miller, Johnston, Hoefer, and Luecke (1954)
	130–200	450–650	250–380	Forbes and Haines (1952)
Child	—	about 1800	about 700	Snyderman *et al.* (1949)
Dog	200–400	1000–2000	500–1000	Axelrod, Lipton, and Elvehjem (1940; 1941)
	250	1300	600	Potter, Axelrod, and Elvehjem (1942)
Fox	250–400	1300–2000	600–1000	Schaefer, Whitehair, and Elvehjem (1947)
Hen	300–350	about 2000	700–850	Hill, Norris, and Scott (1954); Jackson *et al.* (1946); Petersen, Lampman, and Stamberg (1947 *a*, *b*)
Turkey poult	270–350	about 1500	650–850	Bird, Asmundson, Kratzer, and Lepkovsky (1946); Boucher, Patrick, and Knandel (1942); Jukes (1938); Patrick, Darrow, and Morgan (1944)
Chick	275–325	about 1500	650–850	Bethke and Record (1942); Bird *et al.* (1946); Bolton (1944; 1947); Norris *et al.* (1936); Stokstad and Manning (1938)
Duckling	about 300	about 1500	about 750	Fritz, Archer, and Barker (1939)
Rat	about 300	about 1500	about 750	Burch, Bessey, and Lowry (1948); Mills (1943); Nieman and Jansen (1955); Sure (1940); Wagner, Axelrod, Lipton, and Elvehjem (1940)
Mouse (C_{57})	about 400	about 1600	900–950	Fenton and Cowgill (1947)
Lactobacillus casei	250–320	500–550	650–750	Bro-Rasmussen (1955)

[a] From Part II of "The riboflavin requirement of animals and man and associated metabolic relations," by Finn Bro-Rasmussen (1958).
[b] Here and elsewhere in this review when quantities of riboflavin (or protein) are expressed as "per 100 g. feed," the feed is a ration of artificial type containing a sugar, starch, or air-dry cereal, a protein, most often purified casein, oil, minerals, and vitamins, or is a simple pig or poultry ration of much the same sort.
[c] The authorities listed are as given in Bro-Rasmussen (1958).

metabolism of farm animals and, as *Technical Communications,* a review on the feeding of camels, one on amino acids in foods and feedingstuffs and, to be published shortly, reviews of diet in relation to reproduction in sheep and pigs. One or two illustrative points may be described.

The reviews of requirements of man for calcium (Leitch, 1937) and nitrogen (Leitch and Duckworth, 1937) were built to the same pattern: assembly of data from metabolism experiments, classification by age and sex, and elementary statistical analysis to show how retention was related to intake. For calcium there were about 400 studies of adults and about the same number of children. The analysis of the adult material was made on the assumption that no adult ought to retain or lose calcium or nitrogen continuously, and that the amount required for maintenance would be the amount that would give equal chances of gain or loss. The concept seemed, and is, both simple and logical, but was not at first well received in some quarters. Mitchell (1938) condemned the material as a hodge-podge of experiments and the method as “unique” but, as far as calcium was concerned, he could quote estimates in agreement and a year later (Mitchell, 1939) produced confirmation from an assembly of calcium balances analysed by a similar method. Data on nitrogen balance are under review now in the light of modern concepts of biological value.

The Technical Communication on the feeding of camels (Leitch, 1940) arose from a request from an officer of the Somaliland Camel Corps for advice about certain disorders in camels which, he thought, might be due to faulty feeding. Since little had been written about the feeding of camels, an enquiry was sent to correspondents in those parts of the Commonwealth in which camels were still used. From the replies received and what publications could be found on the diet and physiology of camels, a summary was made and a comparison of diets on which camels were said to do well, or not to do well. On analogy with the feed requirements of cattle, about which there was plenty of information, it was at once clear that camels were often underfed, or ill fed, or both. It was then a simple matter to draw up plans for diets adequate by standards for cattle, made up of the feeds commonly given to camels. It was also clear that the camel has a peculiar requirement for salt, which is related to its capacity to withstand deprivation of water; deprivation of salt may cause serious disorder.

This report was received with gratitude by the officer who asked for advice, and, he said, with amusement by his camelmen who thought it funny that advice should come from a remote armchair in a country where camels exist only in zoological gardens. The conditions for being able to make this particular armchair study of a distant field problem (and for giving advice on many equally remote problems that have arisen since) are: (i) the overseas contacts which made the questionnaire a success, and which all Commonwealth Agricultural Bureaux have, and (ii) knowledge of the requirements of cattle and the composition of feedingstuffs, which is part of the everyday stock-in-trade of the bureau, and is kept alive and meaningful by constant contact with the research and farm work of the Rowett Research Institute. The necessary basic information about camel feeding was "retrieved" by questionnaire; interpreta tion and deductions were tested against a large body of related knowledge.

THE CREATIVE REVIEW

So far we have described reviews which assemble, analyse and discuss material mostly from one field of research. But there is another type, the highest and rarest, which takes data from more than one field and shows that they are related and what the relation is. From our own experience we can describe only one, *Food Health and Income* (Orr, 1936). *Food Health and Income* took shape against a background of poverty and food shortage in the house-holds of low-paid workers and the unemployed in the midst of apparent surplusses of agricultural produce. The Ministry of Agriculture and Committees associated with it, economists, and statisticians provided estimates of food supplies and the distribution of the national income, and an assembly of family budgets, woefully small and poorly representative of the population but all that there was at the time. The Bureau had for some time been collecting data on heights and weights of school children from reports of School Medical Officers, and more general information showing, or suggesting, a relation of ill health and high mortality to poverty and poor diet. A preliminary report was prepared, of which the essence was embodied in *Food Health and Income*.

The book condensed and focussed a great deal of thought about agriculture on the one hand and health on the other. Taken along with the researches into diets and requirements in the United States which came from the Bureau of Home Economics of the United States Department of Agriculture under Dr. Hazel K.Stiebeling, it helped to convince agriculture that an apparent surplus of food was in truth a deficit. That conviction led straight on to the proposal of the League of Nations in 1935 to marry health and agriculture, and so to FAO.

Not only so, but there is now in most countries a steady stream of work on the economics of diet, and the survey of supplies and consumption of food has become a national service in many. The World Health Organization and a very large number of students of social medicine are occupied, in parallel, with the economics of health and its relation to diet.

SOME PRACTICAL PROBLEMS

FACILITIES FOR WORK

In order that research reviews may be produced there must, of course, be access to libraries, adequate in their holdings and liberal in the privileges they grant, so that the literature required may be found and borrowed. It must be possible to command a team of skilled or semi-skilled labour to collect and transcribe. There must be expertise in many languages. There must be someone in charge.

The reviews to which reference has been made in this paper involve a knowledge of chemistry, physiology, clinical medicine, embryology, zoology, economics, statistics, animal husbandry, and field work of several kinds. Clearly, the setting up of a centre in which research reviews in biology are to be produced would require either a team of experts or a superior jack-of-all-trades with access to theoretical and practical help from experts. Since whole-time occupation of a team of high-power experts would be extremely expensive, and the work of writing reviews seems not to appeal to many of them, the jack-of-all-trades offers the more likely solution. Such a jack-of-all-trades need not of course be “master of none,” but must be sufficiently knowledgeable about all, not only to be quite sure when and where expertise is needed, but also to put a problem to, and discuss it with, the expert.

It seems likely that such a person, in addition to a natural inclination to such analytical and philosophic reasoning, must have a very broad university education and a number of years spent in at least one branch of biological research. Further, the place where he is to work should be such that he can have day-by-day contact with active research. In that way ideas conceived in academic contemplation may be trimmed and shaped in relation to actual problems.

AWKWARD QUESTIONS

Not only should the centre with its consultant experts be equipped to deal with straightforward assembly and analysis of data, but it may need a sort of corporate imagination to devise methods to deal with awkward questions. There is the problem of providing an approximation when data to give a precise answer do not exist; and that of giving a quick answer when there is not time for the laborious collection of existing data. The following are examples.

Approximation

It is often argued that the cross-sectional picture of growth derived from routine measurements at school of children of different ages is of little value, and the only worth-while study is the longitudinal, measurement of the same children at different ages. There is of course a confusion of ideas there, but ignoring it, a device which was used in our study of secular change in the height and weight of school children goes some way to provide a substitute for a mass longitudinal study. It is, briefly, to compare at successive ages children born in the same year. They are not the same identical children. As records are, they provide at the best groups in which part replacement has occurred; but they have at least lived under the same general conditions of economic life and welfare services. Further study will be required to show in what ways the composite picture may differ from a limited, true longitudinal study.

Substitution

Early in World War II, the question of the policy to be adopted in respect of stock farming, which had been heavily dependent on imported concentrates, called urgently for information on the relative efficiencies with which animals convert feed to food for man. There were a few *ad hoc* classical studies of the feed cost of beef production and some records of milk produced and feed eaten by prize cows at dairy shows, but little else in readily usable form. There were very large numbers of experiments in feeding cattle, pigs and poultry, but to collect and analyse them then would have taken far more time than could be given. A short cut had to be devised. Instead of the results of actual experiments we used the standards on which countless experiments had been based and the amounts of the products budgetted for, and so compared feed costs of the products (Leitch and Godden, 1941). The results were checked against those in the few major studies, and with a small amendment where a standard was revised, proved acceptable and useful in practice.

EPILOGUE

Each of the types of research review discussed above has a part to play in furthering biological thought and investigation. The statistical review and the service review gather together scattered numerical data in logical order, and by so doing open new prospects to theory and practice. The review of concepts helps to dissipate too rigid interpretations of physiological or biological “laws” which may be hindering the formulation of new ideas; it discusses and disposes of sham concepts which merely confuse. The better examples of all kinds of research review are creative by suggesting new hypotheses or converting hypothesis to “law.” The Creative Review, so described, is the highest manifestation of such endeavour because it deliberately sets out to effect a synthesis between phenomena previously unrelated.

Most of the examples chosen to illustrate the argument have arisen in response to a need, more or less urgent, for information that could not be reliably obtained from a superficial survey of the literature. Needs and applications have been in political, social, medical, nutritional, and agricultural fields. Each review can claim to have retrieved and made use of knowledge in danger of disappearance and loss, or to have clarified issues and cleared the way to further research. Indeed, the technique of the research review, by virtue of the assembly and use of scattered records, appears to be unequalled as an instrument for retrieval of buried work. It gives a new value to the small experiment and the single biological observation, and in the analysis may reveal truths which might not be reached in a lifetime of direct investigation.

REFERENCES

ADOLPH, E.F. (1933). *Physiol Revs., 13,* 336.
BEAN, W.B. (1948). *J. Appl. Physiol., 1,* 458.
BLAXTER, K.L. (1950). *Nutrition Abstr. & Revs., 20,* 1.
BOYNE, A.W., AITKEN, F.C., and LEITCH, I. (1957). *Nutrition Abstr. & Revs., 27,* 1.
BOYNE, A.W., and LEITCH, I. (1954). *Nutrition Abstr. & Revs., 24,* 255.
BRO-RASMUSSEN, FINN (1958). *Nutrition Abstr. & Revs., 28,* 1; 369.
CORNER, G.W. (1944). *Ourselves Unborn.* Yale Univ. Press, New Haven.
DARWIN, C. (1859). *The Origin of Species.* John Murray, London.
GRANICK, S. (1951). *Physiol. Revs., 31,* 489.
GREENWOOD, M. (1935). *Epidemics and Crowd Diseases.* Williams and Norgate Ltd., London.
HENDERSON, Y. (1925). *Physiol. Revs., 5,* 131.
HYTTEN, F.E., and DUNCAN, C.L. (1956). *Nutrition Abstr. & Revs., 26,* 855.
HYTTEN, F.E., and LEITCH, I. (1957). 4ème Congrès International de Nutrition, Paris, 1957. *Résumé des Communications,* p. 262.
KEYS, A. (1949). *Nutrition Abstr. & Revs., 19,* 1.
KLEIBER, M. (1947). *Physiol. Revs., 27,* 511.
LAWS, R.M. (1956). *Nature, 178,* 193.
LEITCH, I. (1937). *Nutrition Abstr. & Revs., 6,* 553.
LEITCH, I. (1938). *Nutrition Abstr. & Revs., 8,* 1.
LEITCH, I. (1940). Imperial (now Commonwealth) Bureau of Animal Nutrition. *Technical Communication* No. 13.
LEITCH, I. (1942). *Nutrition Abstr. & Revs., 11,* 509.
LEITCH, I. (1951). *Brit. J. Nutrition, Proc., 5,* 142.
LEITCH, I., and DUCKWORTH, J. (1937). *Nutrition Abstr. & Revs., 7,* 257.
LEITCH, I., and GODDEN, W. (1941). Imperial (now Commonwealth) Bureau of Animal Nutrition. *Technical Communication* No. 14.
LIND, J. (1753). *A Treatise of the Scurvy.* Sands, Murray and Cochran, Edinburgh.
MADDEN, S.C., and WHIPPLE, G.H. (1940). *Physiol. Revs., 20,* 194.

MANERY, J.F. (1954). *Physiol. Revs., 34,* 334.

Medical Research Council Committee on Haemoglobin Surveys (1945). Spec. Kept. Ser. No. 252. H.M.S.O., London.

MITCHELL, H.H. (1938). *Ann. Rev. Biochem., 7,* 353.

MITCHELL, H.H. (1939). *Actualités scientifiques et industrielles, 771, Nutrition 18.* Hermann et Cie., Paris.

MORANT, G.M. (1950). *Proc. Roy. Soc., London, B137,* 443.

ORR, J.B. (1929). *Minerals in Pasture and Their Relation to Animal Nutrition.* H.K. Lewis & Co. Ltd., London.

ORR, J.B. (1936). *Food Health and Income.* Report on a survey of adequacy of diet in relation to income. Macmillan and Co. Ltd., London.

QUENOUILLE, M.H., BOYNE, A.W., FISHER, W.B., and LEITCH, I. (1951). Commonwealth Bureau of Animal Nutrition. *Technical Communication* No. 17.

RUSSELL, F.C. (1944). Commonwealth Bureau of Animal Nutrition. *Technical Communication* No. 15.

RUSSELL, F.C., and DUNCAN, D.L. (1956). Commonwealth Bureau of Animal Nutrition. *Technical Communication* No. 15, 2nd ed.

SHAW, R.R. (1955). 1st International Congress on Documentation of Applied Chemistry, London, November 1955.

SIMON, J. Reports of the Medical Officer to the Privy Council *(passim).*

SIMON, J. (1897). *English Sanitary Institutions,* 2nd ed. Smith, Elder and Co., London.

SOSKIN, S. (1941). *Physiol. Revs., 21,* 140.

THOMPSON, D'ARCY W. (1942). *On Growth and Form,* 2nd ed. Cambridge University Press.

THOMSON, A.M., and DUNCAN, D.L. (1954). *Nutrition Abstr. & Revs., 24,* 1.

Recent Trends in Scientific Documentation in South Asia: Problems of Speed and Coverage

P.SHEEL

ABSTRACT. Outlines the causes underlying delays in dissemination of current scientific literature; reviews recent regional developments in dissemination of scientific information; describes an experiment on speedy dissemination of information based on collection by air mail of the tables of contents of scientific periodicals; discusses the present trend in indexing national or regional scientific papers; and suggests systematic listing of contents of scientific publications at the national level and dissemination of these lists at an international level.

The increased effort in scientific research has resulted in a large increase in the volume of scientific publications: periodicals, special project reports, dissertations, etc. Prompt dissemination of information about the availability of this literature has a very special significance for scientists working in laboratories far removed from the industrially advanced countries. For the active prosecution of research at an increased tempo in India, or other countries of South East Asia, and even perhaps Australia and New Zealand, special problems arise in providing the scientists an effective speedy information service on current literature as also for exhaustive retrospective searches.

THE PROBLEMS

The problems which the scientists and documentalists face in literature searches arise from:

1. Inadequacy of the coverage and the delays in reporting literature in the internationally known abstracting journals.
2. The language problem and the consequent delay in noticing such papers.

P.SHEEL Insdoc, National Physical Laboratory, New Delhi, India.

3. Delays which accrue from the normal postal surface transit of the periodicals from the country of origin to the user institution.

ABSTRACTING JOURNALS

Only a few decades ago the scientists felt that their needs for information on current research were adequately met by organisations publishing the abstracting or indexing periodicals. These are, however, no longer considered adequate.

The large increase in the number of scientific publications and the problem of indexing, etc., arising therefrom have been discussed previously by two Conferences, one convened by the Royal Society of London in 1948 and the second convened by Unesco in 1949. The recommendations made have led to many improvements.

It was estimated in 1934 that the increased number of scientific papers which would have to be indexed would amount to 750,000 (1) while the Unesco Conference estimated this number at 1,850,000 (2) scientific articles. The number of scientific periodicals published in 1952 has been estimated to be of the order of 50,000 (3). Assuming an average of 50 articles per year per journal, an approximate estimate of the number of papers currently published would be about 2.5 million. This is nearly 1 1/2 times that estimated at the time of the International Conference on Science Abstracting (1949).

The number of scientific papers reported in some of the English language abstracting journals in 1945 and 1956 is shown in Table 1.

TABLE 1

	1945			*1956*		
Abstracting journal	*Vol. no.*	*No. of pages*	*No. of abstracts*	*Vol. No.*	*No. of pages*	*No. of abstracts*
1. *Science Abstracts, Section A Physics Abstracts* (monthly)	48	339	3148	59	1,014	9,165
2. *Science Abstracts, Section B Electrical Engineering* (monthly)	48	292	2744	59	566	4,661
3. *Biological Abstracts* (Monthly)	19	2546	23,498	30	3,613	36,080
4. *Excerpta Medica* (15 sections) (monthly)				9–10	11,304	54,537
5. *Chemical Abstracts* (semimonthly)	39	2782	28,655[a]	50	8,768	90,310[a]

[a] Estimated number at 10.3 abstracts per page (4).

It will be seen from Table 1 that significant progress has been made by the international abstracting journals to cope with the increased number of scientific publications—the coverage has increased nearly threefold over the last decade.

LANGUAGE PROBLEM

Scientific papers are published in various languages, the major languages being English, French, German, Italian, Portuguese, Spanish, Russian, Japanese, and Chinese. A smaller proportion of papers is published in other languages of the European countries and also of the South East Asian countries. The proportion of papers appearing in different languages varies from subject to subject. Similarly, the proportion of foreign language papers covered in different abstracting journals varies. The use of different languages adds to the difficulties in prompt reporting of scientific literature. Other problems in this respect have received great thought and a recent publication (5) highlights the pertinent facts.

TIME LAG

TRANSLATION

The time lag between the publication of a paper in a given language and the inclusion of its abstract in the appropriate abstracting journal depends on a number of factors. The average time lag is, however, of some consequence from the point of view of rapid dissemination of information.

In Table 2, data relating to a few papers on mathematics, observed in the *Science Abstracts (A),* 1956, indicate the average time now taken for noticing papers in one branch of science.

It will be observed that it takes about three months for *Science Abstracts (A)* to notice papers in English published in British journals, almost four months for papers in English from American journals, while it takes about six months to notice papers published in a European continental language. The average time taken for arranging the translation of the author abstract or preparing an abstract of a paper by a subject specialist is of the order of 2 to 3 months.

The average delay in noticing papers published in other fields of science, accruing from the language problem, is likely to be of the same order, namely six months. This is borne out by the experience of the editors of the *Bulletin Analytique* (now named *Bulletin Signal etique*), who estimated in 1948 that the average interval between the arrival of the periodical and the publication of its abstract in French was about five months (2).

POSTAL DELAYS

Scientists working in Western laboratories receive the scientific periodicals a week or 10 days after publication, even from across the Atlantic. Closer distances

TABLE 2

Abstract No.	*Original communication, language*	*Date of publication of journal containing the article*	*Date of issue of Science Abstracts (A) in which noticed*	*Delay, months*
		Group 1[a]		
1	German	June 1955	January 1956	7
867	German	Sept–Oct. 1955	February 1956	4
4773	German	January 1956	July 1956	6
5594	German	January 1956	August 1956	7
6290	German	June 1956	September 1956	3
4147	Spanish	Sept–Oct. 1955	June 1956	8
7103	Spanish	May-June 1956	October 1956	4
8546	Hungarian	May 1956	December 1956	7
7098	French	May 1956	October 1956	5
5601	Russian	1955 (translated in Nov. 1955)	August 1956	9
		Group 2[b]		
4146	English (U.S.A.)	February 1956	June 1956	4
4149	English	March 1956	June 1956	3
2620	English	December 1956	April 1957	4
		Group 3[c]		
3436	English (British)	March 1956	May 1956	2
2605	English	January 1956	April 1956	3
2607	English	November 1955	April 1956	5

[a] Average delay, 6 months.
[b] Average delay, 3.7 months.
[c] Average delay 3.3 months.

TABLE 3

Title	*Country and date of publication*	*Postmark issue*	*Postmark receipt*	*Transit time, days*
European journals				
Frequenz	Germany (E) Nov. '57	Berlin 22 Nov. '57	New Delhi 28 Dec. '57	36
Nuovo Cimento	Italy Oct. '57	Bologna 5 Nov. '57	New Delhi 30 Dec. '57	55
Cern	Switzerland Sept. '57	Geneva 25 Oct. '57	New Delhi 3 Jan. '58	70
Acustica	Germany (W) Vol. I, No. 5	Stuttgart 25 Nov. '57	New Delhi 30 Dec. '57	35
Knizhnaya Letopis	U.S.S.R. Nov. '57	Moscow 11 Nov. '57	New Delhi 30 Dec. '57	49
United Kingdom journals				
British Printer	U.K. Dec. '57	London 4 Dec. '57	New Delhi 4 Jan. '58	31
Foundry Trade Journal	U.K. Dec. '57	London 5 Dec. '57	New Delhi 4 Jan. '58	30
Trans. Faraday Society	U.K. Nov. '57	Aberdeen 28 Nov. '57	New Delhi 9 Jan. '58	42
United States journals				
Electrical Engineering	U.S.A. Dec. '57	New York 25 Nov. '57	New Delhi 9 Jan. '58	45
Translation Monthly	U.S.A. Nov. '57	Chicago 19 Nov. '57	New Delhi 9 Jan. '58	51

and quicker transport by sea over the Atlantic are primarily responsible for the quick transit and, therefore, scientists in the West would not normally consider the postal delays as a factor to be reckoned with in the dissemination of scientific information. But for scientists working in laboratories distant from the technologically advanced countries, the time taken in the postal transit of publications becomes an important factor and causes delay in obtaining up-to-date information. Table 3 above gives the actual posting and arrival date in New Delhi of a small random selection of the journals received in December 1957 and January 1958 from various countries.

Table 3 shows that if a scientist receives the periodical, he is in a position to study the new developments reported after a period of about 6–8 weeks. But if he has to depend for information on an abstracting journal, he will have to wait for 20–25 weeks; and for papers published in languages other than English, it may be of the order of 30–40 weeks. Thus practically a year has elapsed since the publication of a paper before the scientist obtains the information!

RECENT DEVELOPMENTS IN DISSEMINATION OF INFORMATION

The average delay in noticing and disseminating information received through the postal channels may not by itself be capable of further improvement. Similarly, keeping in view the requisite coordination between large numbers of operations and agencies, the publication programme of the abstracting journals can be considered as quite reasonable. Yet the delays have led to distinct developments for the improvement of several of the information services.

INTERNATIONAL AGENCIES

The first important development in this direction has been the organisation of good documentation by the international agencies themselves, primarily for their own specific needs, and later on extended to others who apply for information. For example, some specialised agencies of the United Nations such as WHO, FAO, ICAO, and ILO have organised documentation work in their respective specialised field.

NATIONAL DOCUMENTATION CENTRES

The second important development during the post-war period has been the organisation of national documentation centres in various countries. It will not be out of place here to make a brief reference to the abstracting or indexing publications of two national documentation centres: Centre de Documentation du CNRS, Paris, and the Centre di Documentazione Scientifico-Tecnica of the Consiglio Nazionale delle Ricerche, Rome.

FRENCH CENTRE

The French documentation centre publishes monthly, in three parts, an abstracting journal entitled *Bulletin Signaletique*. Each paper indexed is noticed in the original language of publication together with a translation of the title and a short indicative abstract in French. A subject index for the annual volume is not published, thus making retrospective search laborious.

The library of the Centre de Documentation du CNRS receives over 6,000 periodicals. The number of abstracts covering various fields of science in 1956 was 226,190 (Parts I and II).

The broad objective of the publications is to give as wide a coverage as possible for scientists using the French language.

ITALIAN CENTRE

The Italian centre publishes every month a bibliographical bulletin in 15 parts, each part covering literature on a particular field of science. In compiling this bibliographical list the title of the journal is given first and pertinent papers in that journal are then noticed with the usual bibliographical data. Italian and foreign periodicals are covered. The language of the original text is retained. Articles are indexed from 2,500 Italian and foreign journals.

UNESCO ASSOCIATED DOCUMENTATION CENTRES

A number of other countries also have set up national documentation centres to meet their own requirements for documentation services. Unesco has provided technical assistance to a number of countries to establish national or regional scientific documentation centres such as: the Centro del Documentacion Cientifica y Tecnica, Mexico; the Jugoslovenski Centar za Tehnicku i Naucnu Dokumentaciju, Belgrade; the Indian National Scientific Documentation Centre, New Delhi; the Scientific and Technical Documentation Division of the National Research Centre of Egypt, Cairo; and the Pakistan National Scientific Documentation Centre of the Council of Scientific and Industrial Research, Karachi. These have already been established and are functioning.

Lately, Unesco has extended technical assistance to the Union of Burma for the organisation of technical information services in Rangoon. Similarly, technical assistance from Unesco has been requested by the Government of Indonesia and the Government of the Philippines for organising scientific information services. These new centres will start operation in a year or so.

RESPONSIVE DOCUMENTATION

The work and functions of these national documentation centres cover two broad aspects, namely responsive documentation services and active documentation work. Under the former category are included services such as conducting literature searches and compilation of bibliographies; locating and providing a copy of a particular paper; translating or helping in translation of scientific papers from foreign languages into a language with which the scientists of the country are conversant. These are undertaken only on request. The organisation of these services may or may not be backed by a science library, though in some of the countries where either communications are not fast or the holdings of the science libraries are not adequate, it is advantageous to build up a centralised science library. This is of special interest in countries which are technically classified as underdeveloped.

ACTIVE DOCUMENTATION WORK

The second aspect of the work of these national documentation centres, namely active documentation, is the more important one for the present study.

Keeping scientists informed of the latest developments in their special fields of study may be achieved by various means, ranging from the simple circulation of periodicals to the compilation and publication of exhaustive or selective classified lists, with or without abstracts. The method to be adopted by a particular organisation depends on the needs of the scientists served, and the available facilities in staff and finance. A number of centres publish bibliographical lists and a brief description of these publications is given in Appendix I.

SPECIAL FEATURES OF THE BIBLIOGRAPHICAL LISTS

These bibliographical lists are compiled primarily for dissemination of information at the national or, in some cases, at regional level. The contents are selective, for the special needs of the country. In deciding on the layout, featuring, and arrangement of the contents of these lists, consideration is given to the reading habits of the scientists. Since the time that the scientists can devote for the perusal of such lists is limited, appropriate classification schemes are used and the text is arranged in such a manner that the scientist can obtain information on the latest developments in his special field in the shortest possible time. These lists are not intended to replace the existing abstracting journals but rather to supplement them in so far as speedy dissemination of information is concerned. In some cases a translation of the title is provided: thus solving the language problem for the broad objective of speedy dissemination of information about foreign language publications. Such lists are primarily not very suitable for exhaustive retrospective searching, and therefore these lists are not usually provided with annual indexes.

EXPERIMENTS AT INSDOC

The information needs of a fast developing scientific and industrial organisation in India led us at Insdoc to give considerable thought to the problem of speed in dissemination of information on scientific publications. The experiment carried out in Insdoc in this connection will be of interest.

The criteria aimed at by Insdoc for its bibliographical list were: (i) comprehensiveness for the immediate needs of the country as far as facilities permitted; and (ii) speedier supply of information than that available from the existing abstracting or indexing periodicals.

A happy solution of these two fold requirements was proposed by Prof. K.S.Krishnan, F.R.S., Director of the National Physical Laboratory (India). In his view a scientist would normally study regularly about half a dozen scientific periodicals covering his special subject. If he could obtain information about the contents of these periodicals by airmail, through the publishers directly, or even from a central organisation, immediately the format of that particular issue of the journal was finalised, then he would be informed in the shortest possible time of the forthcoming papers and could order a copy of a particular article on microfilm by airmail. In this manner a scientist could actually get the text of the paper much ahead of the receipt of the periodical by surface post.

In view of the difficulties of making arrangements directly with a large number of publishers of the important journals, the possibility was explored for obtaining by airmail copies of the table of contents of a selected number of periodicals from national science libraries. The receipt of this information at short intervals would cut down to the barest minimum the time lag accruing from the surface transit of the periodicals by post. This suggestion was appreciated by the foreign documentation organisations contacted and regular arrangements were entered into with some of them for obtaining by airmail a microfilm copy of the table of contents of about 350 scientific journals, the information about the contents of which was considered important for the immediate needs of the scientists in India. The tables of contents are despatched by the cooperating centres 3 to 4 times a month and reach New Delhi approximately a fortnight after the publication of the journals in Europe, America, or the countries of the Pacific Ocean. The expenditure to Insdoc on this account is of the order of $1,800 per year, i.e., $5.00 per periodical.

The information thus received is disseminated in India through the medium of the *"Insdoc List of Current Scientific Literature"* twice a month, and this bibliographical list is generally in the hands of the scientists in India ahead of the receipt of the periodicals sent through the surface post.

It was decided to present the titles of papers in a classified order, the classification being carried out from the information available in the contents page only. Four years' experience of working by this method has indicated that nearly 85% of the titles are telltale and do not present any difficulty in classification. In another 10% cases, difficulties arise, and the entries are likely to be placed in a wrong category, especially in borderline subjects. A small proportion of papers, about 5%, present difficulties for according an appropriate classification number, and for the present are left unlisted.

The time factor involved in the departmental work of translation, classification and the subsequent operations of composition, featuring, layout, and printing also received consideration, especially from the viewpoint of speedy dissemination. At present Insdoc is able to process—classify, compile, and print—about 2,000 titles in 10–14 days' time after receipt of the microfilms. The information about the contents is therefore in the hands of the scientists in India within a maximum period of about four weeks from the date of publication of the periodical in the Western or Far Eastern countries. In exceptional cases, when the information is delayed even in air-mail transit, a further delay of a week or 10 days becomes inevitable, as the material is then included in the next issue. The procedure adopted in the compilation of *Insdoc List* has been described elsewhere by Reid (7).

A critical evaluation of the *Insdoc List* as an indexing journal has been made by Neelameghan (8), specially in relation to articles on medicine. He was collecting recent references on a medical subject and checked the entries in *Insdoc List* along with those in specialist journals such as the *Current List of Medical Literature, Excerpta Medica, British Abstracts of Medical Sciences,* and *Abstracts of World Medicine*. He found some additional references in the *Insdoc List*. He says:

On analysing these additional references from *Insdoc List* [they] were noted to be of the following types: (1) articles from journals not indexed in other specialist indexing/abstracting services mentioned above and (2) articles from very recent issues of journals yet to be indexed by those services though earlier issues of the same journals have been covered by them. It was also noted that issues of certain journals indexed in the *Insdoc List* had just been received or were expected to be received shortly in the Library.

In a personal communication regarding *Insdoc List,* Dr. A.J.Barnard of the J.T.Baker Chemical Company of U.S.A. has stated:

We have indeed found a number of articles listed [in *Insdoc List*] that we proceeded to examine in connection with certain current projects. Curiously, most of the papers located had appeared in European journals and had been overlooked in other bibliographic sources.

This observation is encouraging since it was made even after the postal transit of the *Insdoc List* from India to the U.S.A.

Coblans (6) has described the *Insdoc List* “as a creditable achievement from the point of view of classification and the effective use of offset printing.”

The experiment on classification and compilation of the *Insdoc List,* it seems to us, has fulfilled reasonably well our expectations in so far as speedy dissemination of information is concerned, but it is still far from the aim of comprehensiveness. There has been an improvement in coverage—from a 16-page issue in 1954 to a 40-page issue in 1957. But this is still a small fraction of all the important scientific literature published throughout the world.

NATIONAL LISTING OF SCIENTIFIC PUBLICATIONS

As mentioned earlier, the national documentation centres have oriented their publication programme to suit the immediate interests of the country or the region they serve. In view of the many difficulties attendant on comprehensive coverage of literature in the abstracting journals, a happy development has been the attempt to index the scientific literature published by the country or the region. Reference has been made in Appendix 1 to the *Boletin* issued by the Mexican Documentation Centre indexing the publications in Latin American countries.

The Unesco Science Cooperation Office at Cairo had been compiling for some years a list of the scientific publications of Egypt and other Middle Eastern countries. These lists are now being published as Part II of the Documentation Bulletin of the National Research Centre of Egypt, Cairo.

The Unesco Science Cooperation Office at New Delhi had been listing, in a U.D.C. classified order, the scientific publications of India, Burma, and Ceylon from 1949 onwards and was publishing it under the title *Bibliography of Scientific Publications of South Asia* as a half yearly publication. From 1955 onwards, Insdoc has been associated with this work. Unesco's Science Cooperation Offices at New Delhi and Djakarta now collect and compile the data on the scientific publications of South and South East Asia (India, Burma, Ceylon, Malaya, Indonesia, Thailand, and the Philippines). Publication work is handled by Insdoc. Attempts are being made to provide as comprehensive a coverage as possible. The publication is now issued under the title *Bibliography of Scientific Publications of South and South East Asia*. It was published as a quarterly journal in 1956 and 1957, but from 1958 onwards it is being published as a monthly. The entries are classified by Colon and U.D.C., as is being done for the *Insdoc List.*

JAPAN SCIENCE REVIEW

The Scientific Information Division, Ministry of Education, Tokyo, in cooperation with other agencies, has commenced publication of the *Japan Science Review*. It is published in four series as a half yearly publication: engineering sciences, biological sciences, medical sciences, and economic sciences. These reviews bring to the notice of the scientists the results of researches reported in Japanese periodicals. The average delay at present in circulating information is over a year.

These national or regional bibliographical lists help considerably in the comprehensive indexing of scientific literature. With improved frequency of publication and lesser time lag in noticing the papers one may hope that the problem of comprehensiveness of scientific reporting will be ameliorated to a substantial extent. The usefulness of the comprehensive listing vis-a-vis extensive abstracting of scientific papers for physicists has been evaluated by Gray (9) in the results of his survey on Physics abstracting.

SUGGESTIONS

The systematic collection of world-wide scientific literature by the organisations which are publishing abstracts, and the subsequent operations of indexing, translation, classification, etc., on a comprehensive basis would in itself be a commendable proposal, but will be beyond the resources of these organisations in the near foreseeable future. Such a plan would, in any case, not eliminate the problem of delays caused by the surface transport of printed matter through the normal postal channels.

The question of organising world-wide listing or abstracting by the formation of regional committees was recommended by Unesco in 1949 (2). A brief account of the progress made in listing national or regional publications in South and South East Asia has already been given in the preceding sections. The satisfactory results of the experiment made by Insdoc in the quick dissemination of scientific information in South Asia, though on a limited scale, offers in our opinion a solution of the difficult problems: comprehensiveness of coverage, language barrier, and speedy dissemination.

In the field of science and technology there are a number of institutions in various countries which are capable of acting as national centres for supply of information about the scientific work of their countries. Such institutions, it is proposed, might be designated as national centres for listing the contents of the scientific publications of the country on a comprehensive basis.

In case the publications are in a language not commonly understood, the national centre could compile them in an another commonly understood language. For example, if the scientific papers published in a Thai journal could be prepared for indexing by a suitable organisation in Thailand and a translation of the titles supplied either in English or French, then it would be possible for the scientists elsewhere to obtain information about the contents of the Thai periodicals which might otherwise possibly be ignored because of the difficulties of language or incomplete dissemination of information.

The national centres are also likely to be equipped with modern photo-duplicating appliances. They may organise the indexing of the national scientific literature by using mechanical aids and copy periodically, say twice a month, on microfilm, the tables of contents of the periodicals published in their countries. Having access to a science library or libraries which receive all the national periodicals and possess microfilming equipment, it should not be beyond the resources of such national centres to arrange this work at a very nominal additional expenditure.

Further, one copy of this microfilmed national list of contents of scientific publications might be despatched by air mail to a central agency under the aegis of an international organisation such as the International Council of Scientific Unions (ICSU) or the United Nations Educational, Scientific and Cultural Organisation (Unesco). This microfilm copy of the national list, received twice a month, could be duplicated in the central agency and a microfilm copy made available, by air mail, to each of the other national centres cooperating in the scheme for compilation of information at the national level. The national centres would then be in a position to disseminate the incoming information much earlier than at present.

A beginning on the lines suggested might be made by microfilming the tables of contents of scientific periodicals only, and this might be expanded at a later date to include the authors' abstracts, wherever available.

A copy of the national lists could also be made available by the central international agency to the editors of abstracting journals, by air mail, so that these journals are also informed of the published scientific work in the shortest possible time.

The improvements in the design and fabrication of mechanical devices for copying, and making multi-copies in limited number on microfilm, and the availability of such machines in the national centres will bring the suggestion within the realm of practical realisation. The cost involved, at the national and the international level, worked out at the prices prevalent in India (*vide* Appendix II), would appear to be moderate.

ADVANTAGES

The suggestion, if accepted and implemented, would cut down the chances of the results of a scientific work completed, published, but not indexed, being ignored. It would contribute considerably to the broad aim of comprehensiveness of literature reporting in the field of science and technology. It would eliminate the delays inherent in collection of scientific literature before it is noticed by the abstracting or indexing journals. It would substantially solve the language problem by providing authors' abstracts or their translation from the lesser known languages into a more widely understood language. And lastly, the use of the microfilm and its transport by air mail would overcome the very long and trying delay arising from the surface transport of scientific literature through the normal postal channels.

REFERENCES

1. *Royal Society Scientific Information Conference, 1948.* Report and papers submitted. Royal Society, London, 1948.
2. *International Conference on Science Abstracting, 1949.* Final report. Unesco, Paris, 1949.
3. W.A.SMITH and F.L.KENT, *Editors. World List of Scientific Periodicals Published in the Years 1900 to 1950,* 3rd edition, Butterworths' Scientific Publications, London, 1952.
4. E.J.CRANE. Chemical Abstracts. *Chemical and Engineering News, 34,* 1238–40 (1956).
5. *Scientific and Technical Translating and Other Aspects of the Language Problem.* Unesco, Paris, 1957.
6. H.COBLANS. United Nations technical assistance and documentation. *American Documentation, 7,* 289–93 (1956).
7. J.B.REID. Insdoc. *Review of Documentation, 22,* 10–12 (1955).
8. A.NEELAMEGHAN. *Insdoc List:* current scientific literature. *IASLIC Bulletin, 1* (2), 37–43 (1956).
9. DWIGHT E.GRAY. Physics abstracting. *American Journal of Physics, 18,* 417–24 (1950).

APPENDIX 1 BRIEF PARTICULARS OF THE BIBLIOGRAPHICAL BULLETINS PUBLISHED BY SOME NATIONAL DOCUMENTATION CENTRES

1. *Boletin* del Centre de Documentacion Cientifica y Tecnica, Mexico is published monthly and lists titles of papers published in 2,500 journals received in the Centre's library. The titles are translated into Spanish. It consists of 5 parts, each can be obtained as a separate.

   Articles published in Latin American countries are indexed with an asterisk mark so as to draw immediate attention to these publications. The number of titles reported in 1956 was 63,807.
2. *Bilten* Dokumentacija Strucne Literature is published monthly by the Yugoslovenski Centar za Tehnicku i Naucnu Dokumentaciju, Belgrade. It is issued in six parts. Each part contains short abstracts in Serbo-Croatian arranged in a U.D.C. classified order. The library receives nearly 2,000 periodicals. Approximately 40,000 abstracts are published.
3. *Insdoc List* of Current Scientific Literature is published from New Delhi twice a month. It indexes titles of papers currently published and classified according to the Colon and U.D.C. schemes. Information about the contents of current periodicals is obtained on microfilm by air mail from a number of centres abroad. Titles in European languages are translated into English. The total number of scientific papers indexed in 1956 was 32,800. No annual index is published.
4. *Documentation Bulletin* of the National Research Centre of Egypt, Cairo, is published monthly, in two parts. In Part I are listed titles of papers from periodicals received in the Library; titles published in European languages are translated into English. Part II lists the contents of the periodicals published in Egypt and other Middle East countries. The titles, if published in the national language, are translated into English. A short indicative abstract is published. The total number of titles indexed in 1956 was 54,478 (Parts I and II).

APPENDIX 2 COST OF MICROFILMING, TWICE A MONTH, TABLES OF CONTENTS OF 100 PERIODICALS AT A NATIONAL CENTRE, AND DUPLICATION INTO 100 COPIES ON POSITIVE MICROFILM BY AN INTERNATIONAL ORGANISATION

	National (one copy)		*International (100 copies)*	
	Titles	*Abstracts*	*Titles*	*Abstracts*
Cost of microfilm	Rs 10	Rs 35	Rs 250	Rs 1,000
Chemicals for developing and printing	Rs 2	Rs 8	Rs 50	Rs 200
Staff time (costs)	Rs 10	Rs 20	Rs 50	Rs 200
Packing and forwarding	Rs 6	Rs 20	Rs 600	Rs 2,500
Total, list/fortnight	Rs 28	Rs 83	Rs 950	Rs 3,900
Total/year	Rs 672	Rs 1,992	Rs 22,800	Rs 93,600
Cost per journal/year, or approx.	Rs 6.7	Rs 19.92	Rs 228	Rs 936
	$ 1.50	$ 4.25	$ 46.00	$ 190.00
			$ 0.5	$ 1.90
Per journal/year/per copy				

ACKNOWLEGEMENT

Acknowledgement is made to the Photographic Section of Insdoc for a trial microfilm of table of contents, with and without authors' abstracts, for estimating the cost involved.

Scientific Documentation in France

J.WYART

The French Scientific Research National Center (CNRS) created, at the beginning of 1940, a documentation service; during the war period, before the invasion of France by the German armies, it was necessary to point out quickly to the laboratories the articles which had appeared in scientific and technical journals that were very difficult to get on account of the war and eventually to provide the laboratories with photographic reproductions of the articles they were interested in.

During the dark war years, while our laboratories lacked most journals and bibliographical reviews such as *Physical Abstracts* and *Chemical Abstracts,* the documentation center was of great help in publishing its bibliographical journal and in providing microfilms from the few journals which reached Paris. It was during those years, when France was isolated, that our scientists became well aware of the sharp necessity to have at hand an efficient documentation service. We had too few scientists and engineers to permit them to lose time and money in solving problems of which the solutions had already been found elsewhere or in building apparatus already out of date.

The documentation center is always continuing its development. At the present time buildings are being constructed which will be large enough for its different services with their 150 full-time workers and 500 part-time workers.

The organization of this center is unusual. I believe it was the first center in the world to associate many different documentation activities which are usually separated, and to include the wide range of all the sciences and all the techniques. Scientists, that is, in the end, users, are responsible for the center's organization and are in charge of it. In France and elsewhere, such a system has often been criticized. Before I describe it in detail, I should like to present some of the principles and motives that have guided us.

J.WYART Faculty of Sciences, University of Paris at the Sorbonne.

THE AIMS OF THE CENTER

The Center of Documentation is available to scientists and applied research engineers in every field from pure mathematics to biology. It makes, day by day, a complete and superficial survey of the content of certain journals. This enables the scientists to mark the papers that interest them, the patents not included, and to make a quicker and better bibliography.

In France the discontinuance of most bibliographical reviews published by scientific societies has helped this new organization and increased funds and more efficiency have been some of the results.

The Center of Documentation was conceived only for scientists and engineers, and not for teachers who need a complete summary written with a critical mind and who must not read the whole article. On the contrary, the scientist is a specialist whose documentary needs are more precise.

The encyclopedic character of our Center has been brought about by the nature of modern scientific work. More and more the scientist has become a specialist, but this does not mean that he is isolated from the other fields of science. Take for instance the case of the crystallographer who specializes in the study of solid state by means of x-rays. He uses in his laboratory the most modern equipment and the best mathematical tools. A score of years ago, this crystallographer would have studied mainly minerals; nowadays, not only mineralogists and geologists, but also chemists, physicists in every field, metallurgists, and biologists will come to him for help in certain problems. In turn, he will need them for his own problems.

Thus scientists are more specialized, but have become, at the same time, more diversified and more in need of each other. The papers they are interested in are now to be found in a great number of periodicals. Not so long ago a physicist needed only to look through a score of journals. For instance, the electron microscope specialist, a so-called physicist, now has to consult a number of important physical, chemical, crystallographical, metallurgical, and biological reviews.

Another aspect of the specialization and of the more universal character of the scientists comes from the drastic change in their relationship with industry. The barriers between pure and applied research have collapsed. It is well known that the most theoretical discovery is followed closely by industrial applications, and Nobel prize laureates come often from industrial laboratories. This shows that the need for documentation is the same for both scientists and engineers.

The Center of Documentation of the CNRS offers the following services: a library, a bulletin of abstracts *(Bulletin Signalétique),* a bibliographical research service, a translation service, a photographic reproduction service.

THE LIBRARY

The Library is only concerned with scientific periodicals from all over the world, 7,000 in all, obtained by subscription or exchange. It also receives complimentary copies from scientific publishers from all over the world. These books are mentioned in the *Bulletin,* after which they are turned over to university libraries or CNRS laboratories.

Owing to its large coverage, the library can provide the most important journal in every field and since it is not a lending library, visitors are always sure to find these journals.

Each year, we add subscriptions, and soon the total number of our collection will reach 15,000. When the new buildings of the Center of Documentation are finished, the Library will have a large reading room for document consultation. Moreover, because of the difficulties encountered in the printing of scientific and technical works, often impossible to reproduce completely and therefore lost for scientific information, the Center of Documentation has been led to offer to scientific workers the opportunity of keeping their manuscripts free of charge in its Library. These works are mentioned in the *Bulletin* and, as for articles, microfilm reproductions may be had on demand.

PUBLICATIONS OF THE CENTER OF DOCUMENTATION

In order to inform scientific workers and specialized documentation services of the content of main journals, the Center of Documentation reproduces their tables of contents, and publishes a 35-mm microfilm monthly review called *Revue mensuelle des Sommaires des principaux périodiques scientifiques et techniques,* which covers about 300 periodicals in such fields as physics, chemistry, and biology.

However, the main publication of the CNRS documentation center remains the bulletin of abstracts, which in a standard format of 21 cm by 27 cm was given the title, in 1956, *Bulletin Signalétique,* instead of *Bulletin Analytique*. The new name fits the concept of the publication better, since essentially it was designed for rapid information in all branches and therefore is concerned only with brief summaries of articles.

The *Bulletin* is divided into three parts:

Part I, Mathematics, Astronomy, Physics, Chemistry, Engineering, Sciences of the Earth

Part II, Biological Sciences, Pathology, Pharmacology, Food Industry, Agriculture

Part III, Philosophy, Religious Sciences, Esthetics, Arts, History of Sciences and Technology, Psychology, Pedagogy, Sociology, Sciences of Languages

Part I and Part II are monthly publications. Part III is a quarterly. Each part of the *Bulletin* contains an Author Index, and each year, three indexes are issued. The three parts of the *Bulletin* are subdivided into sections printed separately.

Part I is divided into eight sections:

Section I, Pure and Applied Mathematics, Mechanics, Mathematical Physics

Section II, Astronomy and Astrophysics, Physics of the Globe

Section III, Physics in general, Acoustics, Thermodynamics, Heat, Optics, Electricity, and Magnetism

Section IV, Corpuscular Physics, Structure of Matter

Section V, General Chemistry and Physical Chemistry

Section VI, Inorganic Chemistry, Organic Chemistry, Applied Chemistry, Metallurgy

Section VII, Engineering

Section VIII, Mineralogy, Petrography, Geology, Paleontology

Part II is divided into four sections:

Section IX, Biochemistry, Biophysics, Pharmacological Sciences, Toxicology

Section X, Microbiology, Immunology

Section XI, Animal Biology, Genetics, Vegetal Biology

Section XII, Agriculture, Phytopharmacology, Food Industries

Part III is divided into two sections:

Sociology

History of Sciences and Technics

On account of the quality of its information, several scientific societies have decided to adopt the *Bulletin,* or part of it, for their own bibliographical research. Such are the International Union of Astronomy, the Geological Society of France, the American Crystallographic Association, and the International Union for History of Sciences, which all subscribe for the sections of the *Bulletin* corresponding to their own branches.

The *Bulletin* differs in its principles from most of the actual bibliographical reviews. The abstracts consist of a title in the original language and its translation in French, followed by a summary expressing the content of the article in several brief sentences in telegraphic style. This abstract is considered sufficient for a specialist to see immediately whether he is interested in reading the entire article, since even with the most complete analysis, the scientist might still wish to consult the original document.

These abstracts must always be only factual. They must not be a reflection of the abstractor's opinion or just a copy of the writer's summary which appears at the head of most articles in periodicals, because authors usually emphasize some new results in such summaries, and leave in the background points of view which may be of great interest to other scientific workers. The abstractor never has to decide whether it is worth while to abstract the article, and the scientific workers can be assured that all the articles in periodicals chosen at the beginning of the year will be analysed.

Not all of the numerous periodicals in our Library are abstracted. This will always be impossible, since a too-voluminous *Bulletin* would be soon unconsultable. For this reason we are reducing our abstracts and are using, as often as possible, only brief sentences. Moreover, this offers the possibility of an easier codification in the near future. In 1957 the *Bulletin* published 163,125 abstracts compared with 22,670 in 1940.

Three different services, working separately, and corresponding to three parts of the *Bulletin* undertake its editing. (1) As soon as the periodicals have been registered in the Library, they are transmitted to the chief editor of each service, who distributes them to the abstractors and supervises their work. Each month a specialist is responsible for the classification of the abstracts within each section of the *Bulletin*. When all the periodicals have been completely abstracted, they are returned to the Library. (2) A common central secretariat assumes the technical preparation of the manuscript and sends it to the printer. (3) The staff of 300 abstractors is composed of full-time specialists and scientists depending on the CNRS, or having an agreement with it. A delay of 5 months is anticipated between the arrival of the periodicals and the publication of the *Bulletin*.

Subscriptions are annual. In 1957, the CNRS Center of Documentation recorded 7,135 subscriptions for the *Bulletin* or its different sections.

The *Bulletin* staff is composed of 3 chief editors, 3 editors, 1 editorial secretary, 83 sections heads (47 for Part I, 23 for Part II, and 13 for Part III), 300 abstractors, and 20 laboratories or organizations collaborating with the *Bulletin*.

BIBLIOGRAPHICAL RESEARCH SERVICE

At the beginning of each month, the *Bulletin* provides carefully classified abstracts, but it is not very convenient for retrospective bibliographical research because it has no detailed subject index, because first it would be difficult to make a good index, and secondly, such an index would soon become unusable. The main difficulty comes from the need to choose two or three words out of the twelve or fifteen ideas that make an abstract. This choice, necessarily somewhat arbitrary, has to be made by good scientists for whom it would be a loss of time. Furthermore, such indexes soon become unusable as they would be much too voluminous and expensive. Most laboratories, in France, make use of the *Chemical Abstracts* index which achieves a kind of perfection in this domain. Every year the index is bigger. The decennial index for the years 1947–1956 contains 19 volumes and 21,500 pages. If nothing changes, it will contain 30 volumes and 30,000 pages in 1966, provided they do not cover more than a few of the 1,000 Russian scientific and technical periodicals and no Chinese at all, which is highly improbable.

This is the reason why, a few years ago, the Center of Documentation started a bibliographical research service to cope with the problem created by the ever growing number of scientific publications. This service will develop greatly in years to come. Every notion appearing in an abstract is translated into code language. This enables a much less voluminous classification; the selection does not have to be made by a specialist any more—a machine does the job. Thus a scientist will be able to subscribe to the bibliography of his specialty, and every month he will receive the abstracts related to it. It will also be possible to provide the past bibliography of any problem.

TRANSLATION SERVICE

The most interesting task of the translation service is to coordinate all the translations made in different French centers. The Center publishes a monthly review of translations and, thanks to a punched card system, can give rapid information as to the existence of a translation and where it can be found. There is also a staff of translators.

PHOTOGRAPHICAL REPRODUCTION SERVICE

This service provides a photographic reproduction of any scientific article, most often in the form of a 35-mm microfilm delivered in 22-cm long strips which correspond to ten pages. A great number of articles can thus be filed easily, and the scientists can have at hand a very specialized library. Every laboratory now has microfilm readers. The Center can provide reproductions of articles published either in any one of the 7,000 periodicals in its own Library or in journals stored in a hundred other Parisian libraries (for instance, articles published before 1940). The service also has a staff of highly qualified librarians who can complete or correct bibliographical references. The number of reproductions made by the Center is high: in 1956, 2,322,300 microfilm pages and 158,000 pages of paper reproduction were delivered, 60% of them to industrial laboratories. International agreements have been made with 17 countries to exchange microfilms. In 1956, the Center received 39,160 pages of microfilm, and sent 85,558 pages abroad.

CONCLUSION

In my description of the different activities of the CNRS in the documentation field, I have tried to emphasize the vital importance of this information service in the fight of all scientists, both in universities and in industry, for better development of the country's economy. This requires close cooperation with other documentation centers in France. Three hundred such centers have been registered and most of them are well organised and efficient in their field. Already, several centers are working in collaboration with the CNRS; 100 of them constitute a “translation pool,” others contribute to the publication of the *Bulletin,* for instance, the CNET (Centre National d'Etudes des Télécommunications); the CDS (le Centre de la Sidérurgie); l'Institut du Pétrole, des Charbonnages de France, de l'Institut Textile de France, de la Fonderie, de la Céramique, des Sociétés Air Liquide, Kodak.

On the international scale, through the ICSU abstracting board (Bureau des résumés analytiques du Conseil des Unions scientifiques internationales), relations have been established with the American, British, German, Russian, and French bibliographical reviews abstracting boards, in the fields of physics, chemistry, and soon biology.

Only international cooperation and mechanisation of the usual means of registration and research will solve the problem of an ever increasing scientific literature all over the world, and allow scientific workers to avail themselves of maximum information.

I think also that the International Scientific Unions can play a prominent part in editing high standard journals and bibliographies. Thus the International Union of Crystallography edits *Acta Crystallographica* with an international editorial board. The material in the papers submitted to referees must be original and must not have been previously published in any language. For retrospective bibliography this Union publishes an annual book entitled *Structure Reports* where the results published in the crystallographic field during the year are presented in normal form and in a critical spirit. It is beyond doubt that, in the near future, such international cooperation in the editing of scientific and bibliographic journals will become more and more important.

Scientific, Technical, and Economic Information in a Research Organization

MAREK CIGÁNIK

The term "scientific, technical, and economic information" (STEI) has become very familiar in our country. No exact definition of this term has been hitherto given, even though many authors have attempted to interpret its meaning. I wish to point out that this composite conception represents the result of two development trends, i.e., on the one hand the effort to improve and to increase the cooperation between special librarians and documentalists and to settle controversies between them, and on the other hand to meet the continuously increasing need for satisfactory information of scientific and engineering workers in their special fields as well as in fields more or less related to their special ones. The latter point appears to be of greater importance than the former. In our country, the essential importance of the latter point was evident in particular in research institutes. Even though some research-type working places have many years of experience, still research institutes and academies of science are, in general, relatively new institutions and thus are more interested in STEI.

The problem of information is not, however, a linear and direct one. It is first of all concerned with ascertaining existing technological procedures and knowledge in a given field, and the possibility of applying knowledge, taken from other fields related more or less to the main field, in the interest of adopting the latest technical progress and eliminating duplicity of work in research. But it is also important to ascertain currently used technics and technology of research started in our plants for comparison with the latest information. Making the results of our own research available for customers, getting an insight into their requirements, and, finally, exchanging information and data by means of publications, lectures, conference, etc., are also topics of importance. Such a view on information differs from those of the librarian and the documentalist, for this activity is beyond the province of a librarian's or a documentalist's work.

MAREK CICÁNIK Cables and Insulating Materials Research Institute, Bratislava, Czechoslovakia.

The division of labour generally practised in today's society made it inevitable for special workers—information engineers—to do the work in the field of information, and to establish individual organisation units designed as scientific, technical, and economic information divisions. This does not mean, however, that the division of labour is in all cases and everywhere strictly maintained, for it is determined by economic rules which must be respected also in the information field. There will be, therefore, cases for which it will be necessary to direct the activity of this division in a different way, since the establishment of separate STEI units is not economical. If we want to solve the information problem with reference to the economical point of view, it is necessary to analyse all related activities and to seek the maximal and minimal limits for the given conditions.

DIVIDING THE STEI'S ACTIVITIES

The STEI divisions comprise a whole scale of simple and intricate activities. Besides information proper, library work, documentation, publishing, bibliography, photoreproduction, etc., are covered. After a detailed analysis in the light of the modern view on information, the conclusion was drawn that there are three types of basic activities comprising all the other activities, namely literature search, literature research, and economic analyses.

Literature search defines an activity steadily aimed at ascertaining sources of information from a determined scientific or manufacturing branch.

Literature research indicates an activity of ascertaining and utilizing scientific and technical experiences discovered in information sources in the course of a literature search, and in addition to literature sources (e.g., samples and products).

Economic studies are thought to illustrate research and technical work effectiveness. The main scope of these analyses is to make it possible to direct technical development towards the greatest possible efficiency.

Literature research is directly connected with research proper by the intermediate of the so-called *indirect research* which deals chiefly with the experimental verification of data taken from literature. Results of literature search, worked out in this way, represent a *verified literature research,* in contradistinction to the non-verified one, and are a necessary basis for proper or *direct theoretical and exploratory investigations*. The results of the verified literature research offer a direct basis, too, for the adoption of up-to-date technics and technology for plants.

The division working in the STEI field activity in our country is called “study division.”

LITERATURE SEARCH

In the literature search activity, we distinguish three stages of work: search proper, treatment of the results of searches, and supply of information (making accessible). Each of these stages comprises several logically connected sub-activities. An overall review of activities concerned with literature search is given in Fig. 1.

It must be emphasised that the most important part of each stage in search proper is to determine the extent of the other activities. In this way, a certain dynamic standpoint has come to the front, for each source is sought and tested from a certain point of view. There is always a main scientific field, to which several borderline subjects have been attached. Sources dealing with the main field will be interesting as a whole regardless of their special application. Sources concerned with borderline subjects will be interesting in their applicability to the main subject and, from a general point of view, by offering a general survey of borderline subjects for research workers.

Thus, for our Institute, technology of cables and insulating materials represents the main topic. We are interested in everything related to this subject. Branches of secondary importance closely related to the main topic are: the whole field of electrical engineering as regards the application of electrical conductors, cables and insulating materials; further the whole field of macro-molecular, chiefly rubber and plastics, chemistry; of varnish chemistry and the related chemical technologies. Measuring techniques and testing practice, mathematics, physics, chemistry, analytical and physical chemistry, machinery and painting technics are considered distantly related to the main topic, principally with reference to the application aspect, i.e., to the cable and insulating material technique.

The general aspect prevails particularly in cases where definite and generally valid methods are concerned, e.g., in analytical chemistry, varnish chemistry, or definite procedures such as drawing and annealing of wires, moulding of plastics, extruding of materials in the plastic state. Thus, when critically examining certain scientific branches in the whole search, three outstanding characteristics are apparent: concreteness, applicability, and generality. The same points of view are applicable also to the main topic, that is, referring to the specialization of the individual research workers, and they are reflected, but in a different way, chiefly when working up the results and making them accessible.

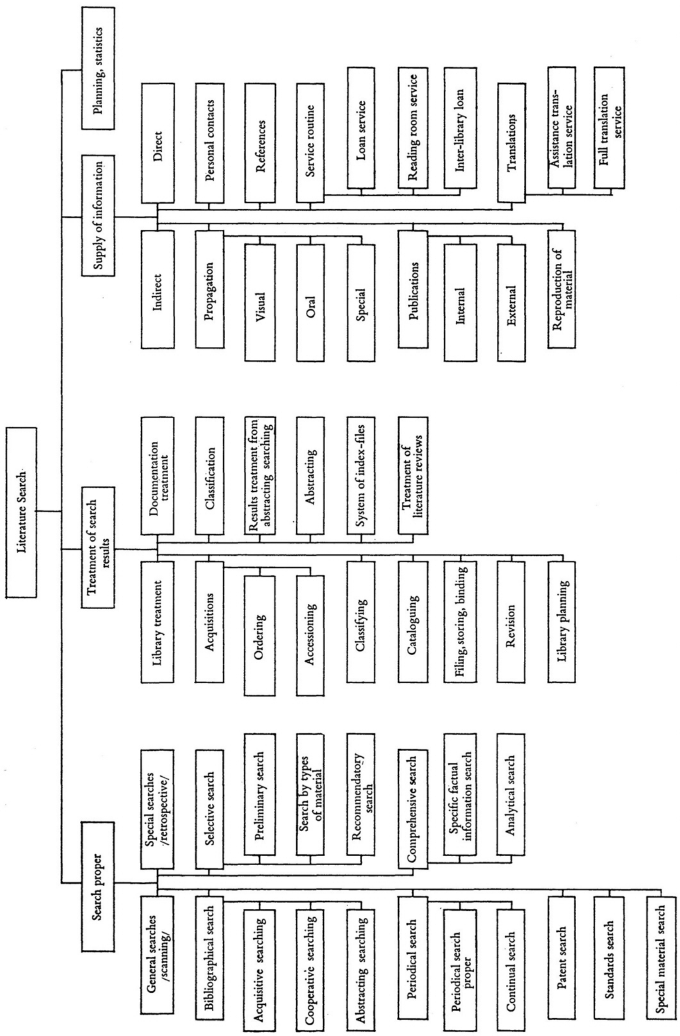


FIGURE 1. Literature search activities.

SUBDIVIDING SEARCH PROPER

Search proper is not restricted to selection of literature on a special topic. That is only a part of the whole search, i.e., search represents, in this sense, a term with new meaning. Search is subdivided into *general* and *special type*.

General search comprises a systematic search in existing literature sources covering a given field as well as fields more or less related to the main field from the three points of view mentioned. Sources ascertained need not be put in practice immediately. As a rule, while performing this part of search, the existing condition or situation is the prevailing factor, but the retrospective standpoint must be considered as well. An essential feature of this part is the continuance of systematic search.

Special search includes ascertaining available data covering a specifically confined subject, such as conductors for X-ray apparatus.

From concepts treated up to this point, it may be concluded that literature search proper may be performed in a high-grade and economic way by a worker having a thorough specialised knowledge not only of the main field but also of those of relative secondary importance. If several scientific fields are covered in the main topic, a further specialization is desirable. A further difficulty lies in the fact that, in addition to a specialization in the related branches, an acquaintance with at least the principal languages is obligatory.

Subdivision of general search

Experience in our country has shown that it is well to subdivide general search according to the types of literature examined. Such a subdivision is instructive, from a theoretical point of view, because, it is necessary, in this case, to ascertain and work out a complete plan of the literature followed and, moreover, it makes possible a further division of labour as well as an effective control of the relative completeness of search performed by a single worker. From Fig. 1 the organization of search is evident, but a brief further illustration seems to be needed.

Bibliographical search. This activity involves searching and ascertaining sources in all types of bibliographies and has a few scopes. First, it has to afford a concrete basis for acquisitions, i.e., it consists of a systematic following of bibliographies, in particular of books and periodicals. In this case a *bibliographic search for acquisition* is concerned. It is impossible to treat either acquisition or selection of general type bibliographies in this paper, for this topic represents a problem common to almost all those who are concerned with technical and often also with all natural science work.

A particular importance is attached to bibliographical search when establishing the plan for ordering periodicals for abstracting and periodical search. In addition to current methods of qualifying and classifying periodicals, or, in other words searches in bibliographies, libraries, and frequency of references gathered from a certain periodical in abstracting journals, we introduced a systematic recording of the frequency of references in a certain periodical on a special card established for each journal. Further details will be mentioned in the section dealing with periodical search.

Within the limits of bibliographic search work, the concept of the so-called *supplementary bibliographic search* or cooperative search is coming to the front. There are two kinds of this search: *(a)* ascertaining the occurrence and availability of a certain publication, e.g., in bibliographies, catalogues of government scientific and related libraries, and in particular of industrial-branch information departments; *(b)* search in specialized bibliographies. Further investigation in these to ascertain sources of specialized searches and completion of sources not found by current methods, from these bibliographies, are practised. [Important bibliographies of this type in our branch are: Underground Systems Reference Book (1931, National Electric Light Association, New York), Underground Systems Reference Book (1957, New York: Edison Electric Institute), Classified Bibliography on Insulated Conductors beginning with 1930 (1954, New York: American Institute of Electrical Engineers), Nau naja literature po dielektrikam (1952, Moskva: Akademija nauk SSSR), Digest of Literature on Dielectrics (from 1936 to date, National Research Council, Washington), Handbuch der Physik, Band XVII Dielektrika (1956, Berlin: Springer-Verlag), Landolt-Börnstein: Zahlenwerte und Funktionen aus Physik, Chemie, Astronomie, Geophysik und Technik, IV. Band-Technik 3. Teil Elektrotechnik (1957, Berlin: Springer-Verlag), Bibliography on Wire (1936–1951, The Iron and Steel Institute).]

Further objectives are search of subject bibliographies and bibliographies as well as search in these bibliographies, and, finally, supplementary search in literature mentioned in high-grade specialized books and papers. Such a search usually gives valuable results.

The last stage of a bibliographical search is the *abstracting bibliographical search*. In our country, this activity is incorrectly called “passive documentation.” Abstracting search has the characteristics of a supplementary search, mainly as an additional periodical search. It consists of following the special abstracting periodicals and *separate cards services,* and of working up sources not involved when performing other types of search.

Periodical search. This activity consists of searching and establishing sources in periodicals, i.e., of searching in scientific and technical journals. This is of utmost importance for the needs of theoretical and fundamental researches. This search results in concrete papers found in a given periodical. Equally valuable are, of course, further data, in particular references to literature, book reviews, advertisements, data concerned with firms, new products, and manufacturers' literature. Thus, the periodical search includes periodical, bibliographical, and trade literature.

A periodical search is a relatively wearisome procedure for the specialized worker and, therefore, a correct choice of periodicals which have to be followed is essential. It is impossible to state the maximum and minimum numbers of periodicals. After a few years' experience, we assembled a list of periodicals actually abstracted (Appendix I). Systematic abstracting of periodicals was started in 1922, when the so-called literary department was established within the library of the Cable Works Co., Ltd., at Bratislava. About 60 foreign periodicals and 5 abstracting journals were abstracted by scientific and technical workers of the concern. Later (in 1947) this number was raised to 130, then to 200, and, in 1955, to 350 periodicals abstracted by workers in the department of documentation. For each periodical, a record card (using the horizontal type Karto system) was kept on which, in three-month intervals, was noted the number of references taken from it, together with a subdivision into concrete, applicable, and general papers. In addition to other well-known criteria, these results are very useful in making a critical examination of a given periodical for our purposes, as well as in preliminary planning of periodicals to be abstracted. Three years' experience with such a type of investigation showed the lack of economy in raising the number of original periodicals followed. For reaching actual sources, a certain optimal number of periodicals is sufficient, i.e., our special requirements are met by following 200 periodicals and 30 abstract journals, which is considerably less than the 350 originally abstracted, the topics being followed very broadly. After this practical evaluation we reduced by one-third not only the number of periodicals followed but also the time needed for search, and raised the number of abstracting journals, thereby broadening the so-called supplementary search. By these measures, we raised the number of sources, while saving one-third the time. The ratio between references taken from original sources and from abstracting journals and cards services is actually 1:1, and the total is 12,000 abstracts a year.

With all searches, but particularly in periodical ones, the ascertaining of sources is related both to the general thematic search plan (main and borderline fields, in general, regardless of whether the source in question is, at this point, interesting) and to the operative research planning (concrete planned tasks and the continuous completion of bibliographies worked out by special search works in the preceding years). This part of search is usually called *continual search,* and this is emphasised especially.

Patent and trade literature search. By this term is meant a search consisting of seeking and establishing sources in patent and trade literature. This represents the most important type of search for industrial and applied research. At this point, it will be analyzed only within the scope of general and continual research. It was apparent that combining the patent and trade literature search was essential. Patent literature, in particular patent applications, precedes all other types of literature, but it is not always realistic. A large dose of phantasy, chiefly in covering, blocking, and securing patents, is incorporated, whereas trade literature is realistic and often gives information on invention realizations.

From the practical and proven economic standpoint, we concluded that following the patent literature of Germany, the United States, Great Britain, and France, and, after careful selection, the manufacturing volumes of the other countries, naturally in addition to Czechoslovak patent literature, seems to be sufficient.

Search is, at our Institute, aimed at official journals and patent abstracting periodicals as well as at patent specifications of selected classes, performed in close cooperation with the National Invention and Standardization Office in Prague. For abstracting patent literature, a special method, described in the section dealing with results of search, was developed.

The objects of trade literature search are specialized journals and house organs mentioned in the section on periodical search, and, in particular, advertisements appearing in these periodical; further, industrial shows, trade literature obtained regularly from firms or centers of trade literature, as well as lectures of individual firm representatives. A prerequisite for the realization of an effective search in trade literature is the availability of a firm directory with a survey on their manufacturing program.

Standards search. This activity is characterized by the ascertaining of existing standards and tentative specifications. Similar conclusions are valid here as in searches in patent literature, but standards literature is delayed as compared with the other types of literature.

Special material search. In this case, a search of bibliographies and bibliographies on research and development reports is made. We follow U.S. Government Research Reports, *Nuclear Science Abstracts,* ERA Reports, Otčoty NIIKP (Naučno-issledovatel̆skij institut kabel̆noj promyšlennosti, Moscow,) and VEI (Vsesojuznyj elektrotechničeskij institut, Moscow), lists of solved and resolved projects of the individual Czechoslovak institutions and others. This literature is less easily available, but it is important.

Another object of this activity is the following of dissertations, symposia, lectures on conferences, as well as of scientific, technical, and instructive films. There is no definite specialization for cable and insulating material technics, but important in this respect are the conferences of CIGRÉ and IEC, World Power Conference, Conference on Electrical Insulating Materials, further meetings of the American Institute of Electrical Engineers, American Society for Testing Materials, Institute of Electrical Engineers, Verein Deutsche Elektrotechniker, and others.

Performance of general search in practice. Between 1922 and 1927, search in the cable concern was accomplished by regular circulation of periodicals among scientific, technical, and commercial workers. The same workers also made the selection of periodicals for search. From 1927 this activity was the responsibility of the director together with top scientific workers of the physical laboratory, which at that time was the research center of the concern. Search was effected on the basis of a plan for systematically abstracting periodicals and patent government publications. Each member of the committee had to follow about 20 periodicals. After the foundation of the Research Institute, the search was effected by the director again, together with a few highly qualified top workers. After the number of periodicals increased, search was transferred to the study department head, and, later, to the specialized group leaders. Until these specialized group leaders acquired sufficient qualifications, search was performed practically by all research workers who received the periodicals in the course of circulation.

The present state of affairs is, in our opinion, the best one. The study department head effects a preliminary search aimed at the up-to-date research tasks, designs papers concerned with these topics, and those which are to be published in the Bulletin. Search is effected immediately after the arrival of the periodicals by mail. In addition, the head of the study department makes the search of the part in patents which cannot be subdivided into individual classes with a definite specialization. The further search, namely a scanning of assigned periodicals is done by the specialized group leader within the study department (engineers), whereas bibliographical acquisitive and cooperative search is done by the librarian-in-chief, and, finally, abstracting and part of supplementary search is done by the individual documentalists with reference to their specialization. A control search, in addition, is effected officially for a definite special field by scientific workers from research, with particular reference to application. This is done in such a way that the study department head, together with the specialized group leaders, indicates significant, interesting papers which must be abstracted within a week. The periodical is afterwards available for readers in the reading room for one month, or, in special cases, it is sent to selected research workers on a “limited” circulation. During availability in the reading room as well as on limited circulation, the periodical should be inspected, in addition to the destined research worker, by all workers interested in the related topic. Every worker who has reviewed the journal, classifies every source in question by “K” (concrete), “a” (application), or “v” (general) according to the principal point of view covered. Further classifications are “obj.” (order!), if a certain source has to be ordered, and, finally, the classification manual and datum are written on the circulation slip put into the periodical. After the journal has been on display in the reading room for an appropriate time, it is returned to the documentalist who, according to his specialty, completes the literature search and performs abstracting work. Uninspected periodicals are also abstracted, but, in those circumstances, on the advice of the study department head.

Pointing out search, in contradistinction to the other activities, in particular the evaluation of results which is closely related to search, is not only of theoretical, but, also of considerable practical importance. Search requires thorough specialized knowledge as well as knowledge of the perspective development trend in a special field. Only when these requirements are met can an effective and economic search be effected without danger of working in the “l'art pour l'art” way. Finally, a well-done general search makes possible an effective division of labour. In Czechoslovakia, where the STEI institutions are relatively recent, it must be realized that a worker, even with university level qualifications, is unable to work out a well-assembled search, and, in this case, the task has to be done by the laboratory and highly qualified workers of the Institute, since neglecting the work concerned with abstracting literature may lead to serious consequences. On the other hand, a highly skilled study department head and specialized group leaders considerably increase the efficiency level of research activity.

SPECIAL SEARCHES-LITERATURE REVIEWING

General type search is characterized by its prevailing practice feature, whereas special search is predominantly retrospective. A finished special search has to answer, in an indicative or complete way, the question of what is known about a specific problem. Special search, therefore, usually was a necessary first stage procedure before the solution of any research, development, or industrial type task. Performance of a special search is, with reference to economy, a first-grade requirement in research and exploring activities. Its methods, however, change depending on the type of basic or applied research covered. According to the solution of a given task that special search has to offer, we distinguish *selective* and *comprehensive special searches*.

Selective special search

This represents a type of search which gives an indicative picture of a certain subject. As a rule, it is restricted to the study department's own findings because information has to be gained within a short time. This kind of search is frequently used in our country at the time the research program is established, when the problem of incorporating the customer's requirement into the plan is decided upon. Selective search has to define the requirement to such an extent that a definite decision can be made. The results of this search are summarized in a so-called *signalizing (preliminary) literature review*. It is often desirable to emphasize, in performing selective search, a certain type of literature such as research reports, patents, and standards. It is usual to designate this type of activity as *selective type searching* (search by types of material). Within a designated type of literature, a complete search, which represents, however, always a selective-type search, may be made. Another point of view appears when classifying a selective search according to the quality of sources; in this case we may speak of a *selective recommended search* (recommendatory search).

Comprehensive search

This search gives a complete picture of the situation in a certain field. Its aim is to ascertain all sources which are related to the topic in question. An absolute degree of completeness of sources, as a rule, can never be attained. Such a requirement is, on the one hand, of no practical value, and on the other hand, it is unrealistic. This lack of practical value is based on the fact that it is unnecessary to quote all sources, which frequently repeat each other. It is necessary, however, to find all sources of importance as well as every source which, compared with other sources, contains, at least in part, novelties. Collected sources are also valuable, of course, even though they do not contain any substantial new contributions, but because they sum up data gathered from widely scattered sources. Such sources may be, after all, compilations of special search results. A requirement of an absolute degree of completeness is unrealistic, for it cannot be attained in practice because of inaccessibility of sources, keeping them in secret, etc.

There are two kinds of comprehensive search depending on the subject covered. When a special detail of a known topic is involved and if it may be resolved definitely by search, we call it a *search for specific factual information*. Thus, for instance, electric properties of polyethylene produced by current industrial procedures have to be ascertained. In this case the search is limited to disclosing concrete data and is neither repeated nor continued; it is, therefore, performed but once. Under these circumstances, it is profitable to quote data from at least two sources, to increase the reliability of data found.

If, on the other hand, a special topic is involved, and the worldwide state of development at a definite time or in a determined interval of time has to be ascertained, an *analytical search* must be performed, because all kinds of literature and all sources available have to be analyzed. Analytical search may be considered of greatest importance because, as a rule, it inevitably recurs in this type of search in which new research tasks, perspective tasks, and important actual research problems, which cannot be provided for within the operative research program, are covered. Analytical search is a basic item of the study department program. A specification that lists accurately not only the requirements of the customer but also data on sources known to the customer is, of necessity, required in such cases.

Practical performance of special search

A specific feature of specialized search is that each kind requires a separate search. The only kind of specialized search which usually is not asked for explicitly, but recommended, is selective recommendatory search.

Special search is generally based on the *requirement list* in addition to search for specific factual information. According to the *requirement list,* the search program is assembled. Although search for specific factual information is not planned, a certain amount of time is alloted to it in accordance with the experience of the previous year. Special search planning is directly related to the research program of the research institute as well as to the technical development program of the enterprise. The research usually represents also a specialized directory center for the whole branch, and participates in the preparations of the technical development plan for the branch in question. A special search program of the documentation development gives, however, no true picture of the research and technical development programs, which can be easily understood. The special search program depends chiefly on the research work and study department organizations, mainly on literary search and research distribution or subdivision. Literature search is usually performed exclusively by the study department, whereas this is only partly true of the literature research which is performed within the study department as well as within the research laboratories. It is further dependent on the kinds of research tasks as well as on the time provided for the solution of these tasks. It is also evident that it ought to be related to general search, especially closely to current search.

The chief problem in special search is its performance in limited time. Work on most tasks starts at the beginning of the year, and it is almost impossible, in the majority of cases, to perform a comprehensive search for several tasks combined with the treatment of results. Accordingly, finishing a concrete search is a matter requiring a lot of time, and the worker charged with the solution of a given task very often wishes to get the results of the search within a short time, so that it is only natural that a controversy arises regarding search velocity and completeness as well as quality requirements. The controversy may be settled easily if a systematic general search of the field in question existed and exists, for the possibility of finishing a special search within a short time and of gaining valuable results increases with the amount of literature found in the files and with the length of the time interval. This requirement is, however, not met in all cases, because the study department's activity is only of recent date. Even within our department, where systematic general search has existed since 1922, the situation looks unfavourable because of the incompleteness of findings and of the lack of literature from 1938 to 1947. For these reasons, it seems necessary to effect a specialized search for a longer time and to plan it in definite stages.

Special search is performed, in general, only by documentalists, including group leaders. Only search for specific factual information can be and also often is effected by the scientific worker himself. It must be mentioned, in addition, that in the study department's activity, for medium- and low-qualification staffs of the institute and for enterprises, selected recommended search is stressed, whereas for research workers, signalling, standards, patents, and analytical searches prevail. An attempt was made to illustrate searches and their relation to research work stages in Table I representing a partly idealized schedule. It must be realized that the schedule mentioned is only for a concrete working place with a defined research task.

During the first research stage, all kinds of general search are dealt with. Special search is closely related in particular to current search which, after results have been worked up, is concluded both by completing the bibliographies and by annual literature reviews related to a concrete narrow topic. An ideal solution in international relations would be for scientific institutes active in the same field to cooperate in publishing such reviews, and thereby a further increase of effectiveness in the field of scientific, technical, and economic information could be attained.

Further stages will be illustrated by means of a concrete example. An airplane manufacturer asks that he be furnished with an aircraft ignition cable stable in a temperature range of −60 to +200°C. Such a cable is not included in the regular manufacturing program and, therefore, the order is transferred to the research institute. This institute entrusts the problem to a specialist, or invites the purchaser to discuss the matter, or, alternatively, requires an accurate definition of the problem, usually in form of answers to prepared questions.

TABLE I *Survey on the topics of literature search and research subdivided according to steps of work in research*

Step of research work	*Topic of literature search or research based on actual problems*	*Stage in STEI Dept.*	*Kind of search and research*	*Users of material*
Initiation of research program: perspective planning and definition of goals	Current awareness in main and more related borderline fields of specialization	General searching	Annual reviews of literature	General management, research management, chief scientists, scientists, engineers
	General survey on less related subjects		General informative and applied search	
	Survey of consumers demand trends and perspective demand programs		General searches	
	Review on basic national economy trends, export, import, and market development		General searches	
	Survey on general questions of management, planning, organisation, and research economy		General searches	
Planning of operation research and development program, exploratory investigations	Selecting projects for research. General survey on the situation in the specific problem field for the applicant and the consumer	Preliminary or signalling searching	Definition of requirements	Research management
	Patent situation in general		Patent selective search	Planning department
	"Classic" and current awareness of trends		Preliminary search	Planning department
	Availability of materials, equipment, etc.		Selective search	Research management
	Review of proposed ideas, market research information, executive investigation		Preliminary economic studies	Leaders of research teams
Accomplishment of research program	Information relating to suggested lines of research, existing and similar type solutions	Analytical searching (Literature studies)	Comprehensive search	Leaders of projects
Literature and theoretical research	Background information for evaluating exploratory research, for theoretical investigations and problems		Comprehensive special material search	Senior scientists and engineers
	Summarizing search results, verification of theoretic data, economic studies, reports of studies		Literature research	Research management

Experimental work and applied research	Information for further applicable procedures and new trends, information for evaluating experimental results and for continuing planning of experimental work	Continual and factual information searching	Continual search Standards search	Leaders of research teams
	Detailed information needed in day-to-day experimental work		Factual information search	Scientists and engineers
Development of research results, semi-plant, pilot plant	Suitability and availability of materials, equipment processes, possibility of application of new methods and technique	Completion searching, patent studies, economic studies	Continual search Complementary search	Leader of research team, development engineers
	Engineering information		Factual information search	Scientists and engineers
	Evaluation of competition and eventual product specifications, tentative specifications, and standards		Special material search	Development engineers
	Ascertaining of economically most suitable solutions and methods		Special economic studies	Research management
	Investigation of the patentability aspects of solutions, internal and external evaluation of reports		Patent search Literature research	Research management General management
Application research full-scale production	Information for gaining data for full-scale production	Elaboration of own research results, continual search, market research	Continuous search	Leaders of research teams and liaison officers
	Information on possible applications and markets		Utilization studies	General and research management
	Background information as to technical practicality, on reasons for customer complaints, plant difficulties, etc.		General search Market research	General and research management

This is the first phase of the second stage involving information on the state of affairs in the problem field and at the purchaser's plant. It is my own opinion that scientific workers from laboratories should not be disturbed repeatedly in an uneconomic way. Search aimed at accurately defining requirements is concluded with the drafting of a report.

After this relatively quick phase the second phase of activity is the comprehensive investigation of the patent field as well as the gaining of basic knowledge on the state of affairs of a topic as given in publications. This in many cases is accomplished in the form of a signalling review. A typical example of a signalling review is given by photographing some known sources, chronologically arranging photocopies and compiling them in box files. A signalling review is studied by the specialist charged with the solution of the task in question or by the leader of the department concerned, and, after a comparison with the actual state of affairs at the purchaser's, a suggested requirement list or, alternatively, a refusal of the requirement is recommended. In a further stage, an accurate definition of the requirements listed is accomplished. In this phase, are gained information and data determined for a preliminary economic evaluation, further data on possibilities of solution in space and time, as well as on the suitability of materials. In this phase, there are often requirements for specific factual information. This stage is concluded by giving a rough suggestion of the method of solution, determining the research team, and signing of the requirement report. The planning department or the *acceptance committee* recommends proceeding further, in accordance with the facts, and decides whether it is necessary to distinguish research stages or only a development stage, and what is given by the nature of the task. The decision has a considerable influence on the extent of the specialized literature search. It is supposed that, under these circumstances, a new problem such as ignition cables with Teflon insulation is involved, where it will be necessary to examine and to verify the properties of Teflon, to resolve extrusion techniques, to suggest a suitable testing method, and so on. A study stage is therefore suggested; the study department works out a comprehensive literature review on a comprehensive literature search; a designated team, of which the literature research worker is a member, effects a literature research; and the study stage is concluded with a study report. When new problems are involved, a study stage is inevitable. There is still another question concerning the division of labour between the study department and research proper. Our experience shows that, if a task with a well-defined scope which can be planned in concrete terms even for a long time in advance is covered, the job of the study department is limited to performing a comprehensive or literature study review. When all the complexities of the problem have to be elucidated in the course of the study stage, the study department work is required to make a literature research also, and this stage is concluded with a study report. In the former case, the documentalist takes part in the research team only in an advisory capacity, whereas in the latter case he is the leading worker, and research workers act as his assistants, in particular when a laboratory verification of study results is desirable. Such a practice has involved applied research tasks. When theoretical and materials research is covered, the situation is different, and the study department concludes its work with a comprehensive search. When a comprehensive search is made only in the third stage, it is not repeated in the course of the following stages. In Table I, comprehensive search is mentioned also in the following stages, but only in case some stage has been left out. Generally, after any comprehensive search, a systematic continual search is started, and a special factual information search is always useful during the solution of a task.

In the fifth stage of research activity, a more detailed economic analysis of the solution has to be effected; in particular the economically most advantageous solution and its dependence on the quality of the solution must be indicated. This work can be performed best by the worker charged with the solution of the task, in cooperation with an economist, the specific worker furnishing the data on the basis of his research and the economist selecting these from the comprehensive literature review or, alternatively, completing them by means of a further special factual information search.

A patent search is made to ascertain the patent aspects of solutions related to a finished comprehensive search. This is concluded with the elaboration of a patent application, or, respectively, with the critical examination of the patentability aspects by a committee within the institute. The prevailing factor in these works is, of course, not search, but the investigation of sources, and, for this reason, patent search concerning novelty belongs to literature research, and thus the accurate quotation of the results of economic analyses has its own place in economic analyses.

In the last stage of research activity, the study department returns to the methods of general search, in particular to its special part, namely to continual search. This phase of the work leads to the publication in the Institute's bulletin of experiences not mentioned in the research report. Thus information from diaries is made available to workers interested further in the subject.

Methods of work used in special search (1).

Some authors, Janicki (2) and Frank (3), distinguish two methods of special search, direct and inverse (direkte und indirekte Methode bei Literaturrecherchen) when describing literature and patent reviews. For the kind of literature review, Frank distinguishes simple literature reviews (einfache Literaturrecherchen) and complete reviews (vollständige Literaturrecherchen). Such a subdivision is by no means complete and makes no difference between search and its results. To this subdivision is roughly correlated special selective and special comprehensive searches. For the selective search, in particular for signalling and type search, the *inverse method* of search is used. Thus, the concept of the signalling review is equivalent to the concept of the simple review, but it differs from the concept of patent, standard, and other *literature surveys*. From the standpoint of source completeness, a selective search is always made by the inverse method even though, from the standpoint of type, the search may be complete.

For the comprehensive search, the *direct method* of search is used, and the concept of a comprehensive review corresponds to the concept of a complete review (vollständige Literaturrecherche), but it is more accurate, for, as has already been pointed out, absolute completeness of search is neither realistic nor practical.

Experience has shown, however, that it is often useful for recommendatory and special factual information searches to use both methods in a certain combination, and for patent literature, it is even necessary. It will be useful, therefore, to introduce a third *combined method* of search by which a certain division of labour for searching activity is made possible. At our Institute, such a division of labour is profitable, because it is possible to shorten the time of search and the treatment of its results. Thus, at the time when the research plan is accurately defined, the group leader uses the inverse method for the signalling search. In a further stage, when the signalling literature review is extended to a comprehensive one, the working program is established in such a way that other members of the study department specializing in standard, patent, and trade literature are called upon. Search is then effected by the team, the leading worker using the direct method and the others the inverse one. It is also possible to realize a division of labour by the participation of outside co-workers active in a related field as well as of further study departments of the branches. This is practised chiefly for patent search concerned with the validity search for patents and also for other kinds of patent searches.

A similar situation exists in the field of special factual information search. This is the most common type of search related to information service in a library and, in our department, this activity is in charge of the chief librarian using the direct method. When, however, he has to solve more complex and intricate problems, he does so with the assistance of a documentation specialist who takes part in the search using the inverse method. Thus search is both accelerated and the reliability of results is increased.

TREATMENT OF SEARCH RESULTS AND ITS RELATION TO LITERATURE RESEARCH AND ECONOMIC STUDIES

For literature search, there are two types of treatment, i.e., library and documentation. With library treatment, the individual sources or source sets are transferred, through the processes of accession, cataloguing, and classification in a signalling way, to the catalogues, and, physically, to the stores. In this part of the procedure a general type classification usually closely related to marking is sufficient. With documentation treatment, at first the point of detailed information and its transformation into bibliographic, indicative, informative, and synthetic type abstracts from all sources including those in a form used for signalling, such as parts of chapters of books and monographs, prevails. A fundamental need for this stage is a thoroughly prepared classifying system such as a special UDC abridged table with a more precise subdivision of the subjects. Abstracts are cumulated in files perpetually absorbing "information energies." When treating lists of references or literature reviews the collective point is again employed. The type of treatment is reflected also in the titles of reference lists. The title of a review of references is determined preferably by the kind of search involved: signalling, current or recommendatory, comprehensive or analytical, patent or standard reviews are distinguished. By extending the title by indicating the kind of review treatment, we may speak of a signalling indicative review, an analytical informative review, a bibliographic list, and so on.

When a study department is not going to take part in the literature research, its work is concluded by finishing an analytical informative review. Such a review represents the result of an analytical search and contains a report on search, the systematic classification of the materials, informative abstracts on each source of importance, bibliographic abstracts on supplementary sources, and, finally, an author index.

The analytical review is sometimes prepared to give a so-called *study review*. Abstracts are omitted. Instead of them, a continuous, comprehensive review with an attached bibliographic list of sources is given. This type of review is advantageous for three reasons: (1) when it represents a survey of the previous year's literature, (2) when it is profitable and appropriate to publish search results, and (3) when the study department is going to take part in the literature research.

I am not going to treat literature research here, even though some types of search, mainly in patents and also economic studies, where market search represents the first stage, are covered. As a practical example of the treatment of general and special search results, a new method of work, developed last year, for the treatment of patent and trade literatures is described. It is based on the author's own suggestion which was verified in practice and tested at VÚKI.

TREATMENT OF PATENT AND TRADE LITERATURES

In this section under the term “patent literature” are covered patent applications, patent specifications, and all types of secondary publications with information on inventions such as the abstracting journals Official Gazette (OG), Abridgements of Patent Specifications (Abridgements), Auszüge aus den Patentanmeldungen (Auszüge), Villa Kartei and others, as well as official patent journals.

With general search, these publications are sources for current and retrospective patent search. Retrospective character is the prevailing factor if, from the very beginning, treatment of patent literature is started.

Trade literature in this section, on the other hand, involves not only manufacturer's literature, like advertisements, prospectuses, and catalogues, but also research reports, lectures of firm representatives, reports on exhibitions, dissertations, etc. This field of literature is the object of a special material search within general search.

It has been emphasized elsewhere that both types of literature mentioned are related to each other in a certain way which has to be considered also when treating search results.

The system of treatment of this type of literature was suggested by the author in 1957 as an innovation proposal at VÚKI, whence it is gradually being introduced into other central working places. The method of work is important where the practically verified, most rational system is given. The various steps are controlled by means of time studies related to the treatment of literature on cable and insulating materials. Four different procedures were examined.

System selected for the treatment of patent literature

Two systems, a less perfect and a more nearly perfect one, are provided for. In the former three types of files, namely two systematic and one accession type file, are used. *The basic systematic file, which, in part, represents a catalogue, is built according to UDC.* It is presumed, of course, that an appropriately compiled abbreviated UDC table for the relative field was prepared. For our use, it was necessary to work out a new schedule for the UDC related to the cable and insulating materials industry suggested as a PE-note to the F.I.D. For this file, a special type of card, namely an A6-size pocket card, open on the upper edge, was suggested. For the practical use of these pocket cards, the reverse side of the pocket should be somewhat shorter, compared with the front side, since it then can be more easily opened. The use of the pocket card eliminates the need for a patent abstract, since, instead, a photocopy of the original abstract with the figure published in the abstracting journal is used. It is no disadvantage if such an abstract is stored in the original language, since it is a more faithful reproduction of the patent itself. In addition, the figure representing an international means of understanding is much more valuable than a translation of a few sentences of the original abstract or a specially prepared abstract. If patents of a country not issuing abstracts are treated, universal abstracting journals such as *Chemical Abstracts, Rubber Abstracts,* and *Technisches Zentralblatt* may be sometimes used or, alternatively, such an abstract has to be prepared by a photographic technique, i.e., by photographing the main claim and the basic figure. It is always better and cheaper than preparing one's own abstract which requires, if it is good, at least one hour's work of a skilled patent engineer or a documentalist.

In addition to the abstract, a microfilm of the patent specification is put into the pocket card. If a certain invention was patented in several countries, all relative microfilms get into one pocket. Storing a microfilm in a pocket is possible in practice. In our field of interest, one specification contains on an average six to eight pages. A six-page microfilm (35-mm size) falls to the bottom of the pocket and so does an eight-page one. Eight- to nine-page microfilms may be advantageously stored crosswise in the pocket to avoid an accumulation of microfilms at the bottom. Microfilms containing more than nine pages are cut and assembled to form a broader tape. It is advantageous to use broad microfilm in sheet form.

Pocket cards are assembled strictly according to the internal UDC tables. To each classification number in the tables corresponds a guide card in the file where, in addition to the UDC number the definition of the classification number, its extent, and cross references are given. The pocket cards contain only the following data:

UDC classification number and first application date	Patent class of the relative country	Patent number
621.315.336.96 “1939.06.14”	Cl. 174–125	U.S. Pat. 2,191,581

Pocket cards between two guide cards are assembled in chronological sequence in such a way that the most recent specifications are always put ahead.

This systematic file serves for a lot of applications. First, it is highly economical, because it combines several operations. Thus it combines a patent or patent application abstract with the original source, accumulates different sources, chiefly equivalent patents of different countries, application, and patent editing, etc. It takes the place of several files and saves space. It also makes possible a quick preparation of complete lists of patent references determined for a certain specialized as well as for a rather general type problem. Filing of relative specifications within the list is decided by the documentalist who uses, as a rule, abstract photocopies and, in doubtful cases, the original, which may be inspected immediately. The great advantage of abstract photocopies is that the original specification does not need to be inspected often, and, in many cases, the abstract in the original language is also useful because it reproduces the patent claim with a higher degree of accuracy.

After search concerned with the treatment of the material for the reference list has been finished, any further work in connection with copying may be omitted, for the list of references is prepared by a photographic procedure. When following this procedure, two aspects may be considered. The abstracts selected may be arranged according to one specification and microfilms according to the second principle. By photographing materials thus assembled, we may get very quickly a list of references with complete original materials on photocopies. Thus, the time interval between the termination of the reference list (reviews) and the acquisition of original sources is reduced to a minimum, which is of considerable importance mainly in research. Reference list microfilms, in addition, make it possible to prepare further copies of the list of references in question for further interested persons at any time, and this results in additional advantages related to economy of space and the possibility of reference list exchanges with other research centers. Preparation of reference list photocopies in itself is profitable, because the user of the list may write his own observations and notes on the blank side of the photostat.

A noteworthy advantage of this file is that after treatment of patents for a certain interval, current lists of references with patent photocopies may be quickly assembled for a particular scientific worker or a particular department, and a complete survey of the situation in the relative specialized patent field may be secured for the workers in question. If a good classification system is established and used, this file is superior to various mechanical procedure systems.

The second systematic file is the so-called evidence card file arranged according to the detailed patent classification scheme of the relative country. Evidence cards are characterized by a collective sheet written for several patents, and, in this case, a cumulative file is meant. Specifications belonging to a certain final subclass are so filed, for instance, German patents of the subclass 21 c 7/01, etc. From the standardization point of view it is advantageous to use A6 for folded A5 size cards. Practically, it is also possible to use but one A6 card containing the same data on the pocket card in the UDC file for each patent, or a pocket card with cumulative photocopies of bibliographic abstracts.

The scope of this systematic file is again a comprehensive one. It makes possible the following of specifications of various countries within the subclasses, and it is used when the novelty of an application has to be ascertained. Its principal use is in procedures concerning patent applications proper. When a foreign patent office objects to certain specifications, it is well to investigate as well other specifications covered in the given subclass. At least, a control of the completeness of reference lists prepared by using the UDC file is made possible, and the file serves as a reclassifying index for the patent classification of various countries and the UDC.

The third type of evidence is the accession file or accession list. It is built according to the patent numbers for each country. For some countries such as the United States and the German Federal Republic, it is appropriate to use a cumulative file for the accession list, because, in the abstracting journals, the accession arrangement of abstracts is maintained, whereas for other countries such as Great Britain it is useful to build a file with cards of A6 size, since the “Abridgements” are arranged according to “Groups” and it is, in the majority of cases, necessary to follow more than one group. This file or list is used as an accession file or list, and makes possible the identification of a specification if the patent number is known. It is advantageous to designate the date of expiration (validity) of the patent, and it may then be used in patent validity searches.

RELATION OF THE SYSTEM TO TRADE LITERATURE

It is not necessary to emphasize which type of relation exists between patents and specified commercial products. However, it is highly effective, for the UDC file of patents, to comprise also abstracts of trade literature. In our system of evidence, there are thematic groups for trade literature, according to which literature is arranged in file boxes. In the central documentation file, there are abstract cards assembled according to the rough abbreviated tables of the related thematic group, but each catalogue is extended to comprise various products classes, the bibliographic abstracts of which are, together with an accurate UDC number, put into the patent UDC file apart from the central documentation file comprising papers and other sources of literature.

IMPROVED SYSTEM

The three types of files mentioned up to this point are, as a rule, enough to secure good patent documentation work. It is possible to improve the system by replacing original sources (or microfilms) for the accession file. The whole system then becomes more expensive, disregarding the fact that the original specifications or their microfilms are not, in all cases, necessary, mainly when a broader following of less related subjects is concerned. The preceding system may be, in addition, fundamentally improved by introducing two further files: an alphabetic *file of firms* (assignees) and an author file of inventors. Costs of establishing such files are relatively low, whereas their effect is considerable.

METHODICAL WORK PROCEDURE

The substantial features of this problem become evident when the most useful working method is used. The last procedure mentioned which was verified in order to establish an improved system makes it possible for four workers— one patent engineer, one photographer, and two typists—to treat 40,000 specifications a year, when a retrospective treatment of specifications is concerned. Such a procedure is, however, not necessarily universally applicable and it has to be modified for the specifications of individual countries. The general principle, however, may be held.

A like procedure may be selected, for instance, for the treatment of U.S. specifications as well as for German patent applications and specifications, whereas a slightly different method of work is appropriate for the specifications of Great Britain and France.

Retrospective treatment of U.S. specifications

(a) The chief preliminary condition for treatment is a general search performed by a specialist reading over each abstract in the *Official Gazette*. The selection may be accelerated by leaving out these classes, where it is quite evident and experienced that no specification involving the relative field of interest will be found. According to the extent of the given field, it is possible to establish with a high degree of accuracy the time necessary for reading over a year's OG materials. Thus, for the cable and insulating materials field in one annual volume 1500 specifications are found, on an average, and the selection requires 40–50 hours. The concrete ascertaining of 1000 specifications requires, on an average, 30 hours. Search is concluded by marking the original abstract in the OG with pencil, and, in addition, by making a note as to acquiring or not microfilm of the original specification, and a bookmark is inserted in the proper place.

(b) OG thus treated are transferred to the typist who writes down the number of the patent and the patent subclass on an A6 size pocket card, e.g., USA 2,187,401 C1.174–121. Pocket cards are numbered in the order of accession so that the advantage of using an abstract journal with abstracts arranged

according to accession numbers may be utilized. Pocket cards are put one upon the other in such a way that numbers written on each cover remain visible. After having treated twenty cards, a cumulative A6 size accession card is written for them. In practice, by copying only the preceding data from the pocket cards a cumulative accession file is built up very quickly. For abstracts marked with a note meaning that a microfilm of the original specification has to be acquired, the pocket card is provided with the stamped letter “M” which is done also in the relative line of the accession card. This arrangement is used as evidence of a quick ordering of original specification microfilms or for loan. It may be done either according to the accession file or to pocket cards arranged later according to patent classes. The whole operation requires twenty hours for 1000 specifications.

(c) After this operation, photographic work is started. It is most useful to microfilm pages of the OG with marked abstracts, but in such a way that the columns with no abstracts in the OG are brought into prominence (by covering). In this case, photocopies of only the desired columns, employing the A5 half-size, may be made from the microfilm. The advantages of this procedure are that less photographic material is required and the cutting operation for photocopies is accelerated. As a rule, however, there is usually more than one abstract in one OG column for the relative subclass. In our field the number is 2 to 3, so that for 1000 abstracts treated only 300–500 microfilms and 300–500 A5 half-size photocopies representing 15 hours' work are made.

(d) A further stage is cutting the individual abstracts, putting the abstracts into pocket cards, and filing the pocket cards according to patent subclasses. At the end of the second stage, pocket cards are arranged in the sequence of accession and, in the same order, abstract photocopies are furnished by the photographer so that cutting and putting into pocket cards is performed sequentially. If it is possible to obtain the original specifications by way of loan (in our country, it is possible for this year, because the Government Office for Inventions and Standardization furnishes all specifications retrospectively until 1945 for appropriate classifications), it is useful in this stage to microfilm and put them into pocket cards together with abstract photographs. In this case, specifications are lent in the order of accession. Otherwise, microfilms of the original specifications have to be ordered according to patent classes to make putting into pocket cards a more continuous process. Since the Government Office for Inventions and Standardization has previous specifications arranged according to the patent classes of the various countries, the latter procedure is also advantageous for securing loans. It is also well for the microfilms of the original specifications to have the patent number visible on the margin. This can be accomplished easily with a Zeiss photographic apparatus used frequently

in our country; for other systems, the patent number may be photographed on a special picture.

Cutting the abstracts, putting them into pocket cards, and filing for 1000 specifications requires 35 hours of work. For microfilming 500 original specifications averaging 6 pages each, 30 hours are required, and for putting them into pocket cards 10 hours are required. It is assumed, here, that for 1000 abstracts 500 original specifications have to be furnished.

(e) The fifth stage consists of classifying abstracts according to a selected system. In our case, a detailed UDC is involved. It is much more advantageous than to use, for example, the German patent classification, because the UDC system is by far superior in degree of thoroughness, systematic work, and accuracy. Thus, the German patent class 21 c 3/01 covers the design of electric conductors and cables and, in this way, patent literature on all conductors and cables is, in practice, to be found in this class. Consequently, it is necessary to select a further subclassification, whereby the uniformity of the system is impaired. The group 21 c 7/01 comprising electric conductors and cables insulation according to the type of the insulating material shows the same characteristics. Within the UDC, there are rough groups for these concepts which may be subdivided to any degree of accuracy, while for the patent classification, such a subdivision is no longer possible.

By arranging the pocket cards according to the patent subclasses of the relative countries, classification proceeds quickly, because several equal or similar type patents are classified at the same time. For each subclass or group, a documentalist skilled in the relative field is appointed, so that correct classification is secured, and 1000 specifications are thus classified within twenty hours. The UDC number is written with a pen both on the pocket card and on the abstract photocopy by the documentalist.

(f) After the classification has been made, additional bibliographical abstracts to be used for the author's file as well as for the systematic file according to the patent classification of various countries should be copied. Copying one bibliographical abstract requires at least 5 minutes, and its filing 1 minute. Such a procedure is possible, but it requires a lot of copying.

It is better by far to eliminate copying and use photography. In concrete terms, the typist is, in the course of work, working with a complete file of patent subclasses. According to the subclasses, she takes out the abstracts from the pocket cards and arranges them in this order by putting the abstracts one upon the other into a special A4 size stencil with transparent tapes placed for fastening. In addition, abstracts are placed one upon the other in such a way that only the heading of the abstract carrying data on the patent number, title, inventors, and firm, application and claims, and, finally, on patent classification are visible. There is place for 40 abstracts on an A4 size cardboard. The whole operation, namely assembling the photographs, which was described briefly, and returning them to the pocket cards after photographing requires 20 hours for 1000 abstracts. From abstracts thus prepared the photographer takes microfilms and an appropriate number of photocopies. This requires 3 hours of his time.

There are three kinds of advantages connected with this work. (1) All data concerning patents treated according to the patent subclasses of the relative countries are made available on the microfilms. (2) Additional interested persons may be informed. (3) If a greater number of interested persons has to be kept informed, photo-offsetting with eventual publication of an information bulletin is possible.

It is further possible to simplify the general treatment, depending on the cooperation of other similar type centers, by exchange of treated patents either according to countries or to fields. Finally, the primary scope of this procedure is attained by gaining cards for additional files as a result of cutting the photocopy. The first photocopy is cut as a whole according to the requirements, and the whole subclass is put into the pocket card. In this way a cumulative systematic file according to the patent subclasses of the individual countries is built. The original pocket cards are transferred to the systematic file built according to the UDC. Additional photocopies are cut for the author's files while, for each firm and each author, one pocket card is put in, and abstracts are gathered for a certain time. After an appropriate number of abstracts has been stored, the abstracts are attached to A6 size file cards. There is place for 10 abstracts on such a card. Cards are selected according to the UDC numbers, and thus is obtained a cumulative file classified according to firms and, in addition, according to fields for each firm. With this method a view of the manufacturing program of the relative firm may be gained and, since this file includes cards on trade literature also, it is possible to recognize the license and patent trends of the firm. In a similar way, measures are taken in the author's file, where, however, arranging cards according to the UDC numbers is omitted.

Cutting 1000 abstracts and filing them requires 35 hours.

Survey on work required for 1000 specifications (net times)

Documentalist	50 hours
Photographer	50 “
Typist	85 “
Cataloger	35 “

It is possible to use a quite similar method for treating German patent ap plications from the journal *Auszüge aus den Patentanmeldungen* which may be combined, in a further stage, with a study of the official patent journal, or, what is more advantageous, with the *card's bibliography,* Villa Kartei, the card of the Villa Kartei confirming that the patent has been granted. Thus, it may be established from the UDC file, for which applications patents were granted, and this is also marked on the cards of the file arranged according to patent classes, the other files not being completed.

For British specifications, the method outlined may be used with the exception that it is impossible to write a cumulative accession card in the second stage, because the individual groups in the "Abridgements" are arranged in the sequence of accession, but it is usually necessary to treat several groups, and this requires a continuous completion of the accession file. The method remains unchanged, however, when the functions of the accession and the patent subclass files are interchanged. Equally well, an additional photocopy may be made during the *photostat assembly stage,* and an accession file may be built by cutting the photocopies, and it is possible again to accumulate abstracts for a certain number series.

A similar procedure may be used also for French and other patents, but, if no abstracting journals exist, an abstract has to be prepared by *photograph assembly.* In the case given, it is useful to request a search in the relative country in the form of a numerical list of references. We proceeded to do so for French patents, and abstracts covering the field of plastics and insulating materials were taken from *Chemical Abstracts, Rubber Abstracts,* and *Chemische Zentralblatt*.

Current treatment of patents

Retrospective treatment of patent literature was the main topic involved when describing the methods of work used. For us, this point is of utmost importance, because following of patents was neglected and now it is urgent to fill quickly the gap that resulted. The system described is most advantageous because, for a high-level documentation center, essentially complete information can be accomplished within two years. It is possible, thus, to treat 80,000 specifications by employing four workers, three of whom are only clerical assistants. The total costs amount to a fraction of that required by any other existing system.

The system indicated has, however, to meet the requirement of a quick and economical treatment of current information on patents. The file system meets these demands very well, but it is necessary to modify the working procedure. That will be shown again with examples:

U.S. *patents*. When treating U.S. patents of the current year, there is no change in the international relation. Only the proportions are different, because the OG is treated regularly, immediately after receipt, so that a great number of patents accumulates and this results in a slight increase in the amount spent for the treatment of a whole year's patents as compared with retrospective treatment. This limitation may be eliminated by performing a proper search and selection immediately after receipt of the individual OG numbers; original specification microfilms are ordered without delay, and the four stages of retrospective treatment are carried out in practice. The fifth and sixth stages are performed once a month, i.e., involving four to five numbers.

German patent applications and patents. The treatment of the patent abstracting journal *Auszüge* follows the schedule mentioned above. Since this journal is printed only on one side, it may be used for cutting. Consequently, photographic work mentioned for retrospective treatment in the third stage may fall off. One number of the periodical may be utilized by several centers, especially those concerned with a different set of problems, so that each center may cut out the abstracts of use to it. The journal can even be used by centers charged with partially overlapping tasks, for it is possible for them to arrange an exchange of subclass photocopies prepared during the sixth stage of retrospective treatment.

Treating the applications of the German Federal Republic does not put an end to work, for the granting of patents has also to be followed for the individual countries. This can be done by following the patent journal or, again, by following the card bibliography Villa Kartei at the documentation center of the Government Office for Inventions and Standardization (SÚVN), and, finally, directly by following the relative subclass original specifications at the SÚVN. When doing this, the cumulative file, arranged according to the patent subclasses, which may be always taken along, becomes important, and the granting of the patent may be indicated on the relative cards. For the sake of completeness the original specification microfilms or cards from the Villa Kartei may be ordered, and these may be, without additional treatment, put into the pocket cards together with application microfilms in the UDC file.

The patents of the German Democratic Republic from the journal *Erfindungs- und Vorschlagswesen* may be treated in a similar way.

Patents of Great Britain and other countries. A similar method of work may also be used for the British “Abridgements” and, as soon as the announced French patent abstracting journal is issued, for French patents. A certain limitation concerned with following these abstracting journals lies in the fact that they are, except the German *Auszüge,* issued with considerable delay. For instance, the British “Abridgements” are being issued two months after the *Official Journal*. In this case, search proper may be performed in the *Official Journal,* and original specification microfilms may be ordered according to it. Pocket cards preferably may be written in advance, and the following search in the "Abridgements" is thereby accelerated.

In our country, there are further possibilities owing to the mutual cooperation of the patent documentation centers and the SÚVN. This cooperation is based on the SÚVN's activity concerned with sending to the individual centers, immediately after arrival at the SÚVN library, the original patent specifications. In this way information on patents is received as much as two months sooner than some abstracting journals. The individual centers usually work out short indicative abstracts for these specifications, and provide them with the relative Czechoslovak patent class numbers which practically correspond to the German patent classification numbers. Concurrently with the original specifications, microfilms or photocopies may be prepared. After treatment, specifications, together with the abstracts, are returned to the SÚVN, which issues abstracting information bulletins for complete fields of knowledge. This cooperation is, at present, in its initial stage, and an economic evaluation has not yet been made. It is evident that, in the form it is practised today, the system is not appropriate because it is economical for the individual centers only when their work is compensated for by the advantage of preparing the microfilm in their own photolaboratories or, alternatively, when using the abstracts from the information bulletin. However, the use of abstracts from the information bulletin involves cutting an appropriate quantity of numbers, attaching them to A6 size cards, and incorporating them into the file. This is a cumbersome operation and by the time these cards are ready for use, foreign abstracting journals are available. Consequently, it is more useful and economical to use the method described in this paper. The cooperation of the agencies is, of course, effective when patents of countries not covered in special patent abstracting journals are treated.

Proposal concerning a more effective utilization of the cooperation with the SÚVN

Even though the method of work suggested in this paper is by far more economical for the treatment of recent patent literature, it shows certain limitations. For the patent center, a complete and high-level patent information service is built. This is very important, for the work in patent policies for a certain field will be ruled from this center. This office will be able to give more comprehensive and accurate information than the SÚVN. It is, however, necessary to make information available appropriately and quickly for workers in research as well as for engineers in manufacturing. SÚVN information bulletins are not meeting these requirements in a satisfactory way, because an indicative abstract usually cannot give the reader sufficient information whether to order and study the relative specification. A fundamental improvement could be attained by the following system. SÚVN would continue to deliver to the patent centers patents included in certain classes. These centers would provide the specifications with German (i.e., in practice, also with Czechoslovak) classification numbers. For themselves, they would prepare specification microfilms which would be put into pocket cards designed for the DC file. For each specification, a very short abstract in one or two sentences is made to give the object of the patent. These abstracts, more properly called short annotations, would be written in columns of the same width as in patent abstracting journals. To differentiate the abstracts, the patent number, in addition to the annotation, would be written on the margin. Abstracts are arranged according to the order of German subclasses. To make the procedure continuous, the typist would arrange the specifications in the order of origin, and afterwards the list of annotations would be written as follows:

Ukončenie obdĺžnikového vlnovodu so špeciálnou úpravou na smerové vyžarovanie pre parabolické antény.	21 c 5/03 Austr. 209.245 Fr. 1,148.363
Spojenie dvoch vlnovodov pomocou okienka z polytetrafluóretylénu alebo polyetylénu.	Austr. 209.032
Sonda na odvádzanie prenášanej elektromagnetickej energie z vlnovodu alebo súosového kábla.	Brit. 787,341
Regulovateľná výhybka vlnovodu umožňujúca prenos mikrovĺn rôznymi vlnovodmi v na-staviteľnom pomere. and so on.	Brit. 787,070

When the list of annotations is finished, specifications provided with the German class numbers are returned to the SÚVN, and a short abstract is written on the specification with a pencil. SÚVN makes the relative specifications available to centers in due time, but only to a selected number, because it is not possible to make an adequate selection for the whole problem of a center, especially when it is scattered over various patent classes. At this point, the center can prepare photocopies, from its microfilms, chiefly of important patents, and send them to the research workers and engineers in plants, according to the thematic plan. Additional work is thereby made economically effective. In a further stage, relative search and treatment of abstracting journals using the indicated method of work are effected. The pocket cards containing data and specification microfilms prepared in cooperation with the SÚVN are already available, and are normally in the operating cycle. When the activity for a determined time interval, such as one month, has to be summarized, the center performs it in the following way. The abstract photocopies are removed from the pocket cards, arranged again on an A4 size cardboard, but in such a way that, in addition to the heading of the abstract, the figure remains and under the figure an annotation is written in the Slovak or the Czech languages that was worked out during the stage of original specification treatment. This is prepared by cutting the annotation list. Abstracts arranged in this way are transferred photographically onto *metal plates* (for off-setting) which serve in the distribution of the *special information bulletin*. This bulletin may be issued either by the patent center or by the SÚVN, the task of the patent center being to prepare only the film for transfer onto the metal plates. Such a bulletin is useful to a certain industrial branch in its concrete specialization, abstracts being much better as a basis for retrospective search and for securing an additional point of view. In the given case, German patent classification, which is at the same time valid for Czechoslovakia, was emphasized. It is also possible to use the International Patent Classification, so that such a bulletin may be of a more universal type. In this case, a sixth point of view is coming in as an addition to the system consisting of five files. Such a system is considerably better and more economical than the system actually used by the SÚVN.

Supplementary data

The whole system is based on microfilms of the original specifications. In practice, in addition to the microfilm, the original specification is available. This is particularly advantageous for U.S. patents obtained in the form of the original specifications. These types have to be clearly distinguished in the file. For this reason, we use a signalling designation with stamped letters. When there is no note whatsoever on the card, only an abstract of the patent or the patent application is available. The mark “M” indicates the availability of an abstract plus the microfilm of the original specification or application. The mark “PAT” indicates the availability of both the abstract and the original literature. In addition, originals are arranged according to the individual countries in the order of accession and in the order of the patent numbers for each country. These are put into special binders. If both the microfilm and the original specification are available, the marks “M,” “PAT” are used.

In addition, gazettes may be lent from the Government libraries, and for U.S. patents, the microfilm edition of the *Official Gazette* supplied by the Micro Photo Inc., 1700 Shaw Avenue, Cleveland 12, Ohio, may be used. This edition has been ordered by our Center from 1940 on.

REFERENCES

1. See the author's report in "Sborník referátů přednesených na konferenci pracovníků závodních technických knihoven a studijných oddělení Ministerstev strojírenství v Brne 30.–31.ledna 1956—lépe využívat poznatků vědy a techniky, 1956, Praha ÚTEIN.
2. JANICKI, W. Über die Kunst des Recherchierens. *Schweiz. Tech. Z. 47,* (1950), čís. 24, str. 383–391.
3. FRANK, O. *Literaturnachweis und Literaturrecherchen,* Stuttgart, Dorotheen-Verlag, 1953.

APPENDIX I LIST OF PERIODICALS AND ABSTRACTING JOURNALS

Abridgements of Patent Specifications Group II,III,IV,V,VIII,XXXVI
A.C.E.C.Revue
Amatérske rádio
American Documentation
Analytical Abstracts
Analytical Chemistry
Angewandte Chemie
Annalen der Physik
Aplikace matematiky
Application and Industry
Archiv der elektrischen Uebertragung
Archiv für Elektrotechnik
Archiwum Elektrotechniki
Aslib Proceedings
ASTM Bulletin
ATM
Auszüge aus den Patentanmeldungen

Beama Journal
Bell Laboratories Record
Bell.Syst.Techn.Journal
Bezpe nost a hygiena práce
Bibliografia elettrotecnica
Bibliotekar
Biuletyn kablowy
Byullete NIIKP
British Plastics
British Plastics Federation Abstracts
Bulletin AIM
Bulletin de la Société Francaise des Électriciens
Bulletin des SEV
Bulletin VÚKI

Cábles et Transmissions
CEIG Berichte
Chemical Abstracts
Chemical and Engineering News
Chemické listy
Chemické zvesti
Chemický průmysl
Chemische Technik
Chemisches Zentralblatt
Chemistry and Industry
CIGRÉ
Communication and Electronics
Corrosion
Corrosion prevention and control
Čtenář
Čsl.časopis pro fysiku

Der Bibliothekar
Deutsche Elektrotechnik
Deutsche Farbenzeitschrift
Digest on Dielectrics
Direct Current
Dokumentation
Doklady AN SSSR
Draht

Ékspress-informaciya élektrotekhnika
Electric Light and Power
Electrical Communication
Electrical Engineering
Electrical Engineering Abstracts
Electrical Manufacturing
Electrical Power Engineer
Electricité
Élektrichestvo
Élektricheskie stantsii
Elektrizitätswirtschaft
Élektrosvyaź
Elektrotechnik
Elektrotechnický obzor

Elektrotechnika
Elektrotechnik und Maschinenbau
Elektrotechnische Zeitschrift Ausg.A
Elektrotechnische Zeitschrift Ausg.B
Énergetik
Energetika (ČSR)
Erdöl und Kohle
Engineering Index (card service)
Erfindungs-und Vorschlagswesen

Farbe und Lack
Felten Guilleaume Rundschau
Frequenz

General Electric Review

Hospodářské noviny

Industrial and Engineering Chemistry
Industrie de Vernice
Informatsionnyĺ byulleteń novo‘inostrannoi nauchno-tekhnicheskoĺ literatury i tekhnicheskikh katalogov zarubezhnykh firm
Informatsionnyi ukazatel standardov
Informatsionno-tekhnicheskii sbornik
Industrie des Plastiques Modernes
Insulation
Izvestiya AN SSSR,otd.khim.nauk

Journal für praktische Chemie
Journal IEE
Journal of American Chemical Society
Journal of Applied Chemistry
Journal of Applied Physics
Journal of Documentation
Journal of Chemical Physics
Journal of Metals
Journal of Polymer Science
Journal of Research of the National Bureau Standards
Journal of Scientific and Industrial Research
Journal of Scientific Instruments
Journal of the American Oil Chemists’Society
Journal of the Inst.of Petroleum
Journal Oil and Colour Chemists Association (London)

Kabeĺnaya tekhnika
Kauchuk i rezina
Kautschuk und Gummi
Khimicheskaya nauka i
Khimicheskaya
Khimiya i khimicheskaya tekhnologiya
Knihovník
Kniźnica
Kolloidnyí zhurnal
Kunststoffe

Lístková bibliografia KVŠT

Magyar Hiradástechnika
Magyar Kémiai Folyóirat
Magyar Kémikusok Lapja
Makromolekulare Chemie

Matematicko-fysikální časopis
Matematicko-přírodov decké rozhledy
Materials and Methods
Materie Plastiche
Mesures & Contrôle Industriel
Microchimica Acta
Mining elect,and mech.Engineer
Modern Plastics

Nachrichtentechnik
Nachrichten für Dokumentation
Nachrichtentechnische Zeitschrift
Nová technika
Nuclear Science Abstracts
Nucleonics

Official Digest of the Federation of Paint and Varnish Production Clubs
Official Gazette

Paint and Varnish Production
Paint Manufacture
Peintures,Pigments,Vernis
Physics Abstracts
Plaste und Kautschuk
Post Office Electrical Engineering Journal
Power
Power Apparatus and Systems
Prilozheniye k zhur.ES “Energetika za rubezhom ”
Proceeding IEE,Pt.A
Proceeding IEE,Pt.B
Proceeding IEE,Pt.C
Production
Promyshlennaya énergetika
Przeglad Elektrotechniczny
Przeglad Telekommunikacyjny
Przemysl Chemiczny

Railway Signaling and Communications
Referativnyí zhurnal-élektrotekhnika
Referativnyí zhurnal-fizika
Referativnyí zhurnal-khimiya
Referativnaya informaciya
Reports Brit.Elect.Res.Assoc.
Resins,Rubbers,Plastics
Revue de la Documentation
Revue générale de l Électricité
Revue Générale du Caoutchouc
Review of current literature relating to the paint.,colour,varnish and allied industry
Rubber Abstracts
Rubber Age
Rubber Age and Synthetics
Rubber Chemistry and Technology
Rubber World

Sciencia Electrica
Schweizer Archiv für angewandte Wissenschaft und Technik
Sdélovací technika
Slaboproudý obzor
Siemens-Zeitschrift

Special Libraries
SPE-Journal
Strojárenská výroba
Strojírenství
Strojnoelektrotechnický časopis

Technik der Lackisolation
Technická práca
Technické noviny (slov.)
Technické noviny (czec.)
Technische Mitteilungen PTT
Technisches Zentralblatt Abt.Elektrotechnik
Technisches Zentralblatt Abt.Maschinenwesen
Transactions IPI
Transactions IRI
Transactions of the South African Institute of Electr.Engineers
Transactions of the Faraday Society

Uspekhi khimii

U.S.Government Research Reports

Vynálezy a normalisace
VDE Fachberichte
Vestnik svyazi
Vestnik elektropromyshlennosti
Villamosság

Westinghouse Engineer
Wire and Wire Products
Wiadomo ci Elektrotechniczne

Zavodskaya laboratorya
Zeitschrift für physikalische Chemie
Zentralblatt der Ungarischen Technik
Zhurnal analitichesko‘kchimii
Zhurnal i teoretichesko‘ fiziki
Zhurnal fizichesko‘khimii
Zhurnal obshche‘khimii
Zhurnal technichesko‘fiziki

SUMMARY OF DISCUSSION

At the outset Dr. Alexander King noted that of five papers in Area 3 dealing with the subject of monographs, compendia, and specialized centers, two papers were on review publications and three dealt with various national arrangements. There were no papers on monographs, compendia, or on specialized information services. He reported the agreement of the Panel, however, to try to retain a trend of continuity by discussing the subject areas proposed and by relying on submitted papers as a point of departure.

REVIEWS

In the matter of reviews it was agreed by members of the Panel that there is a rising demand for quality reviews and great difficulty in getting qualified scientists to write them. The decline of the critical review was noted, but it was the consensus that review literature is going to play an ever-increasing part in the dissemination of the results of research. Dr. King concluded that in a period of intensity of specialization the review article is one of the real methods of securing cross-fertilization, of permitting browsing in related fields, and of stimulating work in interesting directions on which creativity depends.

The review article and the review journal attempt to satisfy information needs by accumulating, digesting, and correlating the current literature in particular fields and giving an indication of the direction which future research might take.

Mr. Dale Baker listed the criteria most quoted as qualities of a good review as: expert writers; critical approach; comprehensiveness; clarity and balance; good bibliography; synopses; and use of tables where suitable. It is important that a writer of reviews decides whether the review is to be a critical approach or simply a summary of current advances.

It was brought out by Mr. Baker that there is a marked difference between fields with respect to the need and use of reviews. As a rule, reviews are in their infancy in new areas of science and only a small portion of the literature is included in them. In well-established fields, for example, in medicine, the volume and importance of reviews have been sufficient to stimulate compilation of bibliographies of reviews. It was suggested that more reviews would be very useful also in the chemical fields.

Mr. Baker stated, "It is becoming more and more essential that reviews should be organized rationally and should be closely geared to the needs of scientists. Some reviews are little more than indexes or annotated bibliographies, whereas others provide more detailed treatment of narrow fields, and attempt a highly critical appraisal of recent work, both individually and collectively with a view to influencing the direction of future research."

The need for a graded set of reviews was stressed. It was pointed out that there should be several levels of reviews based on differing needs relating to scope of coverage and detail. In some fields it is desirable to have frequent summary reviews of current advances and less frequent comprehensive critical reviews. Chemists, for example, engaged entirely in research need comprehensive reviews on specialized topics, supplemented with full and complete bibliographies. Whereas those concerned mainly with lecturing and teaching, ask for broadly based reviews with key references.

Mr. Baker further stated, "Industrial chemists use review material to a considerable extent. Good review material is to be found in private or semi-private house journals of industrial organizations, while in general industrial chemists are less enthusiastic about the annual reviews." He observed that an overwhelming majority of chemists are in favor of confining reviews to review journals, and against having journals publish original work interspersed with review articles.

Dr. Brygoo brought out that in general reviews may be classed according to three forms. In one case the review is effected by the arrangement of abstracts according to a systematic plan. By this method there is obtained an intermediary result between the ordinary form of informative abstract journals and the periodical review which picks up where an annotated bibliography leaves off. A second form is the occasional review which is rarely published with consideration for the actual and present needs of the users but very often as an opportunity for the editor or the writer. Many of these reviews appear to be the result of "scissors and paste" work on the part of a student or a secondary effort on the part of the author.

The third class of reviews (research reviews) is important in several areas of science. In particular, Dr. Brygoo observed that the presentation of statistical data or of concepts included in the enormous amount of original work published each year in biology could be very useful for research workers if such material could be collected and scanned completely and quickly and then adequately presented in reviews.

One of the main difficulties in obtaining good reviews is simply that not enough qualified people are prepared or are able to give the time necessary to produce them. Scientists who have specialized in library and information work are comparatively unknown. This is less true in the industrial field but in the more rarefied field of academic research the library chemist does not have the status nor is he thought to be so essential as the laboratory chemist. Panel members agreed this attitude is one which if it prevails, might well cause a breakdown in the whole system of information storage and retrieval, for it is essential, if the most efficient methods of organizing scientific literature in a scientific way are to be utilized, that a proportion of first-class scientists be encouraged to specialize in documentation and library information techniques.

It was further stated that the best review is written by a specialist in the field and not by a special kind of scientist. During discussion of the difficulty of getting competent people to write reviews it was suggested that younger research workers might be utilized for this purpose and be forced, as it were, to learn their own subject by reading the literature and producing for one or two years the basic material for reviews. In answer to the criticism that the young research worker does not possess the necessary background to be adequately critical, it was proposed that reviews should be written by teams which would include the young worker and a specialist in the field.

There was consideration as to whether a central documentation service with specialists and computing machines tied in with telecommunication equipment could alleviate the need for literature reviews. There was general agreement, however, that the development of machine documentation and central information services would not stem the increasing demand for review articles.

MONOGRAPHS

In introducing the discussion on monograph publication, on which no papers were submitted, Dr. King pointed out that prompt publication of original papers by present day scientific journals in most countries has reduced the need for monographs. But the monograph has great advantages in that it often includes a number of pertinent papers and it is a useful vehicle sometimes for the longer papers which cannot be easily published in a journal.

Several definitions of a monograph were presented. Dr. Pietsch and Dr. Urquhart observed that in European minds the historic monograph is a form of publication which brings together all pertinent things about one question. It was agreed that this type of monograph has deteriorated in importance but that this historic trend may have been reversed with regard to monographs published in the Soviet Union.

Dr. King concluded that there is a whole group of non-serial publications which may be generally labeled monographs and which include the traditional type of monograph, the original contribution published separately, and, perhaps, the modern technical report. On this latter point there was considerable question as to whether the technical report deserves special consideration as a new medium of publication or whether the report is merely a collection of preliminary observations and data of which the most important will appear later in journal form. There was some opinion that technical reports do not merit abstracting if it can be assumed those which contain important information will later appear as journal articles.

COMPENDIA

In the absence of papers dealing with scientific compendia, Dr. Pietsch described the basic character of the compendium and reviewed the role it has played from its inception to the present. He pointed out that while abstracts and special reviews are considered essential, questions are raised frequently with regard to the usefulness of compendia, perhaps even their right to exist.

Two criticisms are directed at the conception and usefulness of the handbook. There is first of all a critical and increasing time gap in the publication of information which must be reviewed against the background of the ever-increasing flow of scientific data. Secondly, the preparation of the handbook is so costly that the price of the compendium becomes prohibitive for the individual scientist.

In favor of the handbook there is the conviction on the part of many that the handbook presents the only means of overcoming the deep compartmentalization of knowledge in a given discipline. The compendium differs from the abstract journal and the review in that it is a long-term and comprehensive record of a given science.

With regard to the field of chemistry in which compendia have been prominent, it is clear that almost from the very beginning of modern chemistry there was a need for publications reviewing the progress made in these fields. It is significant that the annual reviews did not completely satisfy the requirements of the practicing chemists and scientists. During the period when the entire knowledge was recorded in no more than a dozen periodicals, several handbooks started to appear. Thus, in the early years of the handbook it became obvious that this type of publication would not claim to be up to date, and that the time lag between the generation of knowledge and its recording in the handbooks was unavoidable.

To illustrate the situation further, Dr. Pietsch cited as an example the preparation of the *Gmelin's Handbuch der Anorganischen Chemie* and the attempts of the Gmelin Institute to overcome the restrictions of a classical compendium. All work processes of the Institute are being continually reviewed for efficiency. An average of 19 literature references per page is cited at a cost of $6.15 for each reference. Every effort is made to streamline the operations and to render them more efficient and more economical through the introduction of modern documentation techniques. This has resulted in substantially increasing output but a significant time lag remains.

Dr. Pietsch observed that two hundred years of chemistry up to 1950, reported in about 60,000 pages, will be brought into the eighth edition which is scheduled for complete publication by 1966. At the same time the Gmelin Institute is preparing the material in the form of card files which will be recorded on magnetic tape and thus readily available through type-written lists for every subject.

The need for specialized compilations of data seems to be growing almost daily. Mr. Hilsenrath cited the revision and updating of the International Critical Tables as an example of the enormous growth of data collections. He concluded that the ten volumes of critical tables would have to be expanded into hundreds of volumes in order to do the job adequate to the present needs that the International Critical Tables did about 25 years ago.

Professor Bernal stated his conviction that compendia are absolutely necessary but that we are in danger of destroying compendia by dating them. He offered as a means of overcoming the delay factor that essential data tables should be compiled in an abbreviated form “in handbooks which you could actually hold in your hands.”

Sir Alfred Egerton proposed that the way to tackle the vital question of compendia and particularly of critical tables is by bringing out series of critical data in special subjects and by keeping that going as far as the advance takes place. This could be done by publishing bodies under competition to provide the scientist the best working material.

In summary, there was strong opinion in favor of compendia, both of the type of large compilations of data as well as the type of specialist publications in a critical sense, field by field. Dr. King concluded either type of compendium is economically an enormous job involving large resources. “It seems to me another case of what we require from a basic world service for science, and I don't think it is a subject on which we can afford very much duplication, although we wish and we require an essential critical approach, especially on the data side.”

SPECIALIZED INFORMATION CENTERS

Any consideration of the need for specialized information centers raises immediately the question: Why does the scientist need aids to information access other than those already discussed—abstracts, reviews, monographs, and compendia?

The answer appears to lie in the different levels of need on the part of the research scientist. The use of abstracts, reviews, compendia and other aids of a similar nature in the past have provided sufficient information for the conduct of an investigation.

Increasingly, however, the specialized information center represents a new level of comprehension demanded by the intensification of research activity in many fields and by the importance of achieving efficiency in the use of information.

The last few decades have been marked by the growth and creation of centers of this kind all over the world, much more in relation to the applied sciences than to the fundamental sciences. Two general types of specialized information centers have appeared: One organized to provide service over narrow segments of particular disciplines, and the other devoted to a national point of view comprehensive in scope and centralized in administration.

The three papers in Area 3 relating to specialized information centers all illustrate the national point of view. In his review of the paper by Dr. Sheel, Mr. Green raised questions as to alternative ways of speedily providing the minimum of information to scientists throughout the world. Radio and television were suggested as an improvement over airmail dissemination. Two interesting points raised by Mr. Ciganik's paper were discussed. One was the utilization of patent information in two phases of their operation; the other had to do with the proper framing of the technical inquiry. In this last instance, an inquiry is assigned to a specialist who either has a face to face discussion with the inquirer to make sure the question is accurately understood, or a series of carefully phrased questions is sent out for answering.

Dr. Kotani described the work of national and regional information centers in Japan. The work and function of these centers are twofold. The one which is called responsive documentation services covers research, duplication and so forth, made on request of the users. The other one described as active documentation work consists of publishing bibliographies and lists of publications of science literature. The publication activity is cross-filed into two categories. In order to keep scientists informed of the latest development in their special field of study, most of the centers publish cross-filed lists of literature. In some cases these lists are very comprehensive and contain abstracts and can be regarded as abstracted journals. In other cases, however, the list is merely a list of titles and author names of papers. The well known INSDOC list is an example. The Japan Information Center of Science and Technology recently established is now publishing cross-filed lists of papers in several branches of pure and applied sciences, and in these lists a simple annotation of contents is given in addition to titles and author names. In order that these lists can be very useful to the scientists, delay must be avoided as far as possible.

The Japan Information Center of Science and Technology is receiving current issues of 400 journals by air from the United States and Europe. These copies are regularly sent from New York, USA, and from Europe by air cargo, and air cargo is much less expensive than airmail. Dr. Kotani stressed that the system of sending journals as air cargo should be internationally organized, in order that scientists in different parts of the world may be able to have access through regional centers to new publications without undue delay. This is a program of no small significance to scientists in Asian countries.

Dr. Kotani pointed out another important function of these national or regional centers is found in the compilation and publication of bibliographies of scientific periodicals appearing in these regions. He referred to an example in regional bibliographical reprints, from Yugoslavia, from Mexico, from India, and from Egypt containing literature from African and Near Asian regions. In Japan the Japan Science Review is published. Most of these regional or national bibliographic journals are written in languages used in wider areas of the world, mostly in English.

Dr. Kotani concluded: “I think English is now coming to be regarded as an international language in the science field. Personally, I think that the adoption of one definite language, an international language in kind, is highly desirable. This can be done in such a way that scientists not only will tend to write papers in this common language but will talk in the common language on the occasions of international conferences and meetings to exchange ideas.

“When once this ideal has been attained, the regional or national bibliographies will be much easier tasks to do. However, in the present state of development in which most papers are written in local languages, national or regional bibliographies are very useful and almost indispensable for making scientific information produced in these areas available to the scientists of all the world.”

Mr. Fry noted the essential characteristics of a specialized information center are completeness and depth in search, and promptness in service. The center is expected to have a rather complete grasp of the recorded information in a particular area, depending upon the degree of intensification of research it is expected to serve. The special points of view often peculiar to particular fields also argue in favor of the specialized information center.

A recent survey, tabulation of forty specialized information centers revealed the many features of service in highly specialized projects in narrow subject fields.

1 The principal information search facility most commonly includes a catalog, abstract file, punched cards, extract file, and tape recordings.

2 The information file normally does not contain a collection of library literature.
3 There is a concentration on providing unpublished data and on distributing basic documents.
4 A majority of centers prepare review reports and compilations.
5 The following means of disseminating information were found to be most prevalent:

(a) Preparation of abstracts, accession lists or bibliographies.
(b) Detailed answers to technical inquiries.
(c) Provision of technical consultation.
(d) Provision of special laboratory services.

Mr. Fry emphasized any attempt to describe the need for specialized information centers must have regard to the actual and potential contributions of available and proposed services at various levels, for example:

1 International cooperative ventures.
2 Government-sponsored scientific and technical information centers.
3 Information centers in technical and professional societies.
4 Information centers in private research institutions and universities.
5 Company information centers.
6 Information centers operated by individuals.

Mr. Clark stressed the importance of giving more attention to the provision of adequate literature services to the applied sciences. He pointed out there is in the United States no centralized source and no centralized documentation system dealing with the large field of technology.

In this connection the very considerable efforts of the Institute of Radio Engineers to provide its own publication and documentation services were described. The Institute publishes a journal of several thousand pages per year, and also the 25 professional groups comprising the Institute each publishes a set of transactions, usually four volumes per year. In addition, to utilize the maximum communication by personal contact, the Institute has organized a system of about 50 symposia and annual meetings.

The Institute, in order to provide abstracts and index service to the most important publications, publishes the abstracts compiled by the radio organization of the Department of Scientific and Industrial Research in London. Mr. Clark also referred to the important contribution of the technical press in various countries which publishes a great deal of original material.

In commenting on the inadequacies of present day scientific and technical documentation organizations, Mr. Clark observed that most concern is in the technical field: "We can see from the papers that have been presented at this conference the trend is in two directions: The establishment of specialized centers to meet narrow, detailed needs and the establishment of great national centers which include provision for specialized needs but cover the whole literature of science, and particularly technology."

On the question of centralized national systems or decentralized national systems, or separate specialized information centers, Professor Mikhailov stressed the advantages of the centralized approach to information services. He compared the position of scientists seeking information to that of "a gold prospector faced with the Himalayan jungle and the rock formations of the modern libraries." He reported his conviction that the most effective way of assisting the scientists was in the creation of a large central institute which in its activities would embrace the entire output of national as well as international scientific experience accumulated so far. He noted the individual scientific trends and disciplines of today are organically interconnected and have their interests in the adjacent areas. It is therefore difficult to draw lines between them in order to avoid duplication of material and data. Professor Mikhailov reported it has been the experience of the All-Union Institute of Scientific and Technical Information over a five-year period that this complex problem can only be resolved successfully within the system of a centralized information institute which has as its assignment the embracing of the entire range of scientific knowledge. In this way he stated it is possible to reach the individual scientific disciplines with knowledge and information taken from the adjacent fields. He cited as an example that useful information for a physicist may be derived from the disciplines of biology, chemistry, metallurgy and other sources, which is not possible if the scientists are concentrating only on one narrow specific field.

"Thus in creating an abstract journal, let us say in the area of chemistry and other sciences, within the system of the centralized information institute it becomes possible with more and greater confidence to evaluate the fullness of coverage of all the primary sources. And finally, another and the most important reason is the concentration within a centralized institute of the specialized personnel and the possibility of using all the data and all the advances of machine engineering available, which are only possible in a centralized service."

Professor Mikhailov stated that in Russia a monograph is understood to be a scientific work which concentrates in full and develops the aspects and facets of a given specific scientific discipline. On this basis the type of literature describing various technological processes for various developments in certain areas could not be classified as monographic literature.

Professor Mikhailov announced that the All-Union Institute is planning to publish beginning in 1959 “monographic reviews or reviews in the form of monographs which will be known as ‘The Total Development of Science.' The volumes that we intend to publish encompass the total development of science, and, divided into various disciplines, will be issued at different times. Of course, the time will depend on the state of the contemporary development of this particular science.

“We realize the difficulties and hardships in the way of publishing such volumes devoted to particular disciplines. The overriding difficulty is the availability of special personnel qualified to write such volumes.”

Dr. Majewski stated that in Poland a practical solution to the problem has been found through a type of organization permitting the use of specialists attached to the several specialized information centers for particular topics. “It seems to us that this system collects the products of the research workers and puts this to the attention of those who can use it.”

Professor Bernal observed that the problem of world communication already had been posed to the Conference. “I would say that we ought to consider that we are building a service system, actually, for world science, and we ought to lay down the foundations for it at this meeting, using as nuclei the ganglias that we have already heard from the Soviet Union, from the United States and other places.”

Mr. Farradane stated his belief that specialized information services to be of real value must be more creative. He suggested there is “too much documentation and recorded information and not enough knowledge, real ideas and creative work essential to the future development of information services.”

On the question of a monograph representing the author's personal evaluation of the literature, Professor Mikhailov stated that at the All-Union Institute: “We consider that the amount of work and energy and thought and scientific skill that goes into a monograph must not limit itself only to the viewpoint of one author, but must embrace all the accumulated knowledge and experience bearing upon this particular field.”

In summarizing discussions of this Panel, Dr. King stated his belief that certain minimum services are needed more and more on a world basis. “We have reached the stage when no country, not even the largest, can be self-sufficient in science. In the course of the next few decades we shall see a rise of scientific activity in every part of the world…it will happen that gradually the number of contributions made to science will be approximately equal in proportion to the population wherever it may be…. Consequently, if we restrict our contacts and knowledge to what is done in our own country or published in our own language, we shall miss a very high proportion of what is really valuable and thus limit intellectual and economic growth. We are at the threshold of the period when we must consider these matters on a world basis.

"We are at present going through a very useful and interesting stage of evaluating the needs of scientists and of industry, discussing and evolving techniques of documentation. We shall have to proceed then to consideration of how to provide the maximum service with the minimum of bureaucratic obstruction. This will not be easy as bureaucracy is inherent in large-scale operations. But it is possible without breaking down the traditional activities and responsibilities; and it is necessary, if we are to prevent expensive and unproductive duplication of services and publications in contrast to the creative duplication of scientific effort in the laboratory."

BERNARD M.FRY, *Rapporteur*
ALEXANDER KING, *Discussion Panel Chairman*

AREA 4

ORGANIZATION OF INFORMATION FOR STORAGE AND SEARCH

Comparative characteristics
of existing systems

AREA ORGANIZATION

Authors of Papers

Members of Discussion Panel

Chairman: ERIC DE GROLIER, Centre Français d'Échanges et de Documentation Techniques, Milan, Italy

C.J.DE HAAN, The Netherlands Patent Office, The Hague, The Netherlands

C.D.GULL, Computer Department, General Electric Company, Washington, D.C.

R.M.HAYES, Business Systems Department, The Magnavox Company, Los Angeles, Calif.

Gerald Jahoda, Technical Information Division, Esso Research and Engineering Company, Linden, N.J.

J.W.KUIPERS, Itek Corporation, Boston, Mass.

JACQUES SAMAIN, Departement Filmorex, Société Ceramique et Mecanique, Paris, France

CLAIRE K.SCHULTZ, Remington Rand Univac Division, Sperry Rand Corporation, Philadelphia, Pa.

HANS SELYE, Institute of Experimental Medicine, University of Montreal, Montreal, Canada

B.C.VICKERY, Akers Research Laboratories, Imperial Chemical Industries, Ltd., The Frythe, Welwyn, Herts, England

EUGENE WALL, Engineering Service Division, E.I. du Pont de Nemours & Co., Inc., Wilmington, Del.

PROPOSED SCOPE OF AREA 4

THE DISSEMINATION OF information to serve the needs of the scientific community, the institutions of learning, and the public requires orderly procedures and systems that will make scientific communications widely available, that will promote awareness of what is available, and that will provide prompt access to any desired information.[1]

Many different systems are in existence for organizing, storing, and searching scientific information. These systems range in size from the small collections maintained by an individual to the very largest libraries; in degree of organization from those arranged only by random order of accession to those involving highly specialized classification, indexing and coding schemes; and in methodology from simple hand operations through large manual systems such as card catalogs to special collections that are partially mechanized. All of these systems are effective to some degree for some particular local use.

It is one of the major objectives of this Conference to encourage the preparation of papers that will meet a high standard of scientific method in the study of existing and proposed systems for the processing, storage, and search of scientific information. Descriptions of procedures should be sufficiently detailed to enable a competent specialist to repeat the procedure and obtain the same results; and disclosures and evaluations of methods, equipment, and systems should be given objectively in sufficient detail to permit both understanding and appraisal by others competent in the field.

Similarly, the comparative analysis of the relative advantages and disadvantages of various systems for different uses should be based upon objective consideration of the efficiency, capabilities and limitations, and costs of such systems considered in the full range of operations from accession through storage to making the desired information available.

Such objective accounts of the characteristics of existing systems (including any mechanisms used) will provide the necessary basis for the design of more effective reference tools, for the development of new and improved equipment to facilitate information storage and search, for the definition of performance requirements, and for the realization of improved systems.

[1] The proposed scope of the Conference Arec, as shown here, was prepared during the Spring and Summer of 1956 and provided to all potential contributors as a guide to the aims of the Conference.

This area of the Conference may be directed toward the following problems:

1 Preparation of objective accounts of systems that have been tried, including specific data on efficiency, capabilities, limitations, and costs for all phases of storage and search operations, and including operating characteristics of machines, storage media, and methods used.
2 Comparison of systems in terms of costs, efficiency, and functional characteristics where more than one system has been used for the same type application.
3 Development of criteria for the comparison of systems. These criteria should incorporate both user and administrative requirements, including relationships to other aspects of information services.

It is recognized that seldom will it be possible to obtain actual comparative studies involving two or more types of information systems used for the same collection. The Conference Committee urges that interested individuals undertake such studies wherever possible. However, it is thought that carefully detailed studies of individual systems can be used as the basis for comparative analysis of the advantages and disadvantages of these various systems. These studies should include common factors pertinent to comparative performance evaluation.

Such common factors will probably include, but not be limited to, the following:

1 Characteristics of the collection: size; rate of growth; whether general or special purpose; number of fields covered; variety of fields covered; extent of coverage in field (s); whether acceptance of new items is controlled (e.g. technical library vs. ASTIA), and so forth.
2 Accessioning and cataloguing: skill and training required of personnel who receive and identify new items; kind and number of accession registers or listings maintained; kind and number of listings, abstracts, etc. resulting from identification that are made available to users; minimum rate at which identification must proceed to prevent increasing backlog; tolerance for delay in making new items available to users, and so forth.
3 Classification and indexing: skills and training required of personnel who classify and/or index items; type of classification system used and whether hierarchical; depth of analysis; whether classification and index terms indicate relationships between subject content elements; open-

endedness of system; extent to which different levels of analysis are used; number of classification categories; average number of items per category; range of items per category; provisions for increasing or decreasing number of categories in terms of numbers of items falling in a category; time required to determine proper category of a typical item; cost per item of classification and identification processes; similar considerations for indexing, and so forth.

4 Search entries: skills and training required of personnel who encode search entries, of personnel who inscribe search entries; procedure used in encoding; type of coding system; number of search terms available; average number of search terms used per item; number of character spaces required per search term and whether system uses dedicated space; space available per item for pertinent search terms; extent of redundancy in code terms; whether search entries indicate structure and interrelationships between search elements; extent of cross-referencing; whether cross-references require or permit multiple entries of item information; time required to encode all search entries for an item; cost of encoding search entries for an item; tolerance for errors in encoding; techniques used in inscribing search entries; time to inscribe; costs of inscribing; tolerance for errors in inscribing, and so forth.

5 Storage of items: total storage capacity required for collection; density of storage (items per cubic foot, items per unit of storage medium); characteristics of storage media; whether storage homogeneous as to media; whether storage compartmentalized; whether storage locations or locator symbols are related to classification or indexing system; cost of storage; time to store an item, and so forth.

6 Storage of search entries: whether combined with storage of items; total storage capacity required; density; media used; whether media erasable and reusable; whether compartmentalized; cost; time to store an entry, and so forth.

7 Formulation of search questions: whether special knowledge of classification, indexing, and coding schemes is required to formulate questions; kind of people who interrogate system; tools available to assist searcher in framing his question; whether question must be encoded or inscribed before processing; time required to formulate questions; time required to encode or inscribe questions, and so forth.

8 Search procedures: technique(s) used in conducting search; whether search must be of all entries or of entries in most probable fractional sections; access time; search rate per unit time; time to complete a

typical search; whether more than one search may be conducted at one time; whether more than one section of search entries can be searched at one time; typical number of searches required to be conducted in a given time; time required to set up a new search; delay allowable before searches are conducted; average number of entries searched before search is complete; cost of conducting a search; variety and number of search types; level of recognition logic required; limit on number of search elements that can be compounded into single question; what portion of search entry must be searched in order to make selection and whether screening devices used; ease of broadening or narrowing question as search results are inspected; extent to which search procedure is self-organizing; extent of intercommunication between searchers and system during progress of search, and so forth.

9 Search results: output of the search (listings of locator keys, listings of titles, abstracts, information items, facsimiles of information items); whether results are adequate; average number of omissions; tolerance for omissions; “browsability;” whether results are relevant in terms of average number of extra selections and tolerance for extra selections; resolving power; whether association reference trails are provided; provisions for repetition of search; time and costs of getting from output to desired end results, and so forth.

10 Retrieval from storage: techniques for retrieval of items from storage; form of item as retrieved (original, copy, excerpt); access time; cost of retrieval; cost of reproduction if any; whether original physically removed; tolerance for delays in accessibility due to prior retrieval or reproduction; time to restore removed items to storage; cost of restoring removed items to proper place; number of items to be retrieved in a given time, and so forth.

11 Revision of system: ease of adding new items to collection or of replacing obsolete items; extent to which reclassification becomes necessary and at what time intervals; effort required to effect partial or complete reclassifications; effort required to re-encode entries for items reclassified; time required to make changes in classification, indexing, and coding schemes; cost of making such changes; effort required to insert new or revised search entries into system; effort required to reorganize compartmentalization or fractionation of storage; capabilities required of personnel or mechanisms in putting changes into effect; tolerance for use of dual systems during reorganization or reclassification; extent to which machines used for search and/or retrieval maintain statistics on actual use; extent to which storage system is self-organizing, and so forth.

Since not all of these factors are of equal weight, and since some are more pertinent in some situations than in others, it is requested that studies of existing systems should also reflect which factors are most critical and which are less essential in the particular situation reported.

The analysis of existing systems in terms of specific characteristics and of factors common to information storage and search systems in general should thus yield criteria by which systems can be evaluated in the future as to effectiveness in any given situation.

Recommendation for a Questionnaire Survey

The Advisory Group recommends that a questionnaire survey be made, under the auspices of the Conference, to gather from system users pertinent data on the factors that are outlined in our proposed statement of the area, in advance of the Conference itself. Such a survey would assure coverage of systems that might otherwise be overlooked in the solicitation of papers, as well as providing basic material of important interest to the discussion panel for Area 4. Such material should also be useful to Areas 5 and 6.[2]

Recommendations for Papers

It is recommended that papers on the description of storage and search systems currently in use should include consideration of both those using manual methods and those using mechanized techniques, and should reflect the differing needs of different types of users—from the individual scholar to the largest library center.

Specific papers on mechanized systems that are actually in use at the time of the preparation of the paper, papers on the comparative evaluation of two or more systems, and papers on the development of criteria for comparative evaluation are especially invited.

[2] Unfortunately it has not been possible to organize and carry out such a survey prior to the Conference.

Conventional and Inverted Grouping of Codes for Chemical Data

EUGENE MILLER, DELBERT BALLARD, JOHN KINGSTON, and MORTIMER TAUBE

In an earlier paper (1) a theoretical demonstration of the basic similarity of all storage and retrieval devices was attempted. The ability to describe storage and retrieval systems in terms of a common set of variables with overall cost, measured in capital investment and human time, as the major distinguishing characteristic, was one of the first results of the general theory:

One of the first insights we gain from the attempt to formulate a generalized storage and retrieval theory which is analogous to communication theory is that the theory must be invariant with respect to the physical form of diverse systems. Edge-punched cards, interior-punched cards, Microcards, Uniterm cards, Batten cards, magnetic tapes, or even conventional alphabetical or classed catalogs may differ from one another in cost, ease of storage, ease of retrieval, size, complexity, etc., but they are alike with respect to basic potentialities for handling different types of literature searches. For equal amounts of coding space all systems can enter an equal amount of information and for an equal number of needles, reading heads or electronic circuitry, all physical systems deliver the identical product for any search (2).

This general conclusion was further elaborated in another paper (3); but, until now, it has not been possible to present an empirical and concrete demonstration of the validity of this theoretical approach. Such a demonstration would consist of a cost comparison of two different systems, each of which contained identical stores of information and delivered the same response to the same question. If two operating systems are compared, it is usually difficult to insure that the indexing is equally adequate or that background and peripheral service functions of the systems do not affect their operation and cost.

In the "Conference Plans and Criteria for Papers" issued by the Planning Committee of the International Conference on Scientific Information, this problem was recognized and it was suggested that "It appears that persons and groups who are engaged in the development of systems embodying new

EUGENE MILLER, DELBERT BALLARD, JOHN KINGSTON, and MORTIMER TAUBE Documentation, Inc., Washington, D.C.

principles for organizing subject matter could effectively compare their systems by applying them to a common set of documents."

It is not always easy to isolate a set of documents; and it is still more difficult to insure the same competence of indexers to perform the analyses. It is fortunate, therefore, that a report of a recent innovation by the Office of Research and Development of the Patent Office (4) describes an isolated group of patents indexed in a manner which makes it practicable to set up the same information (the same group of patents) in another storage and retrieval system and obtain a direct comparison of costs for the two systems.

The systems to be compared are a Matrex system using inverted code groups and an International Business Machine punched card system using conventional code groups. Before proceeding with the comparison, some preliminary remarks about devices are in order.

The Matrex device, Fig. 1, is a species of the Batten system, like the Peek-a-boo system, using superimposed aperture cards and light penetration to detect the matching of codes. It differs from most Batten systems in its use of a drill for multiple entering rather than a punch which enters codes one card at a time.

The Patent Office system utilizes a standard IBM keypunch and a Census Bureau multicolumn sorter which is a modified IBM 101 electronic statistical

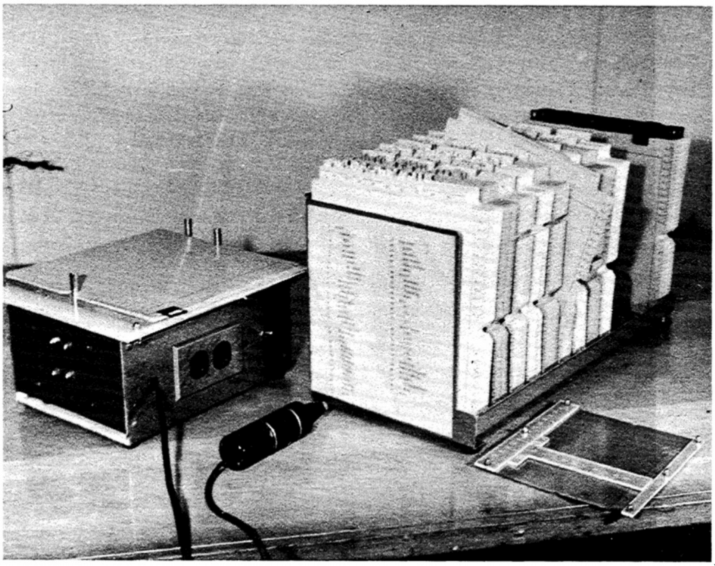


FIGURE 1.

machine equipped with dial switches (instead of a plug board) for setting up the search elements. In this paper, we shall consider this machine roughly equivalent to the IBM 101.

A summary of the Patent Office system

The Patent Office system covers 2350 steroid patents which include 900 duplicates or cross-references leaving 1450 patents to be coded. A steroid compound, the subject of a steroid patent, is a complicated chemical structure. The basic nucleus of a steroid has 22 points at which different chemical groups or substituents can be attached. According to the Patent Office chemists, a particular steroid compound is describable by locating any one of 24 descriptors at any one of 22 positions of the steroid nucleus. Figure 2 (5) gives both the names of the descriptors and a table of positions. The compound illustrated has a "=" at the 4th, 9th, and 17th position; or allo at the 1st, 3rd, 5th, 11th, 17th position; COOH or COOR at the 20th position; etc. In addition to the 24 descriptors requiring position designations, 60 descriptors are given which characterize a steroid without reference to any position. These are indicated in Fig. 3. (6) According to the Patent Office Report:

The punched card is roughly divided into two sections. Columns 1 to 48 are reserved for recording the descriptors for substituents associated with the steroid nucleus at a designated location thereof. The second portion, columns 60 to 69 is reserved for general descriptors not identifiable with any particular position on the nucleus (7).

The first 48 columns are divided into 24 two-column fields for the 24 substituents, and each two-column field provides 22 positions (with 2 left over) for the positions at which the substituents are associated with the nucleus. Although there are two unused holes in each of the 24 fields, no use is made of them for indicating the general presence of a particular substituent without reference to position. Apparently this information is not used in coding or in a search.

Figure 4 (8) illustrates the punched card corresponding to the compound shown in Fig. 2. The card can be read quite easily. Note that "=" (columns 1 and 2) has been punched in the 4th, 9th, and 17th position, etc.

Since this paper is concerned with the cost of entering and retrieving a given quantity of information from two storage and retrieval systems, it is assumed that the indexing of the patent, the determination of suitable descriptors, etc., are entirely adequate. The problem, then, is reduced to entering the descriptors in a Matrex system and searching the system by any descriptor or combination of them.

	0																				11	12
		1									2											
=	22	1	2	3	(4)	5	6	7	8	(9)	10	11	12	13	14	15	16	(17)	18	19	20	21
		3									4											
α or allo	22	(1)	2	(3)	4	(5)	6	7	8	9	10	(11)	12	13	14	15	16	(17)	18	19	20	21
		5									6											
-C≡C-	22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
		7									8											
CH_3	22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
		9									10											
CN	22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
		11									12											
COOH or COOR	22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	(20)	21
		13									14											
-C- sub	22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
		15									16											
-H	22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
		17									18											
NH_2 or N <	22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
		19									20											
O H	22	1	2	(3)	4	5	6	7	8	9	10	(11)	12	13	14	15	16	(17)	18	19	20	(21)
		21									22											
=O	22	1	2	(3)	4	5	6	7	8	9	10	(11)	12	13	14	15	16	17	18	19	(20)	21
		23									24											
-Se-	22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
		25									26											
-S-R	22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
		27									28											
Hal	22	1	(2)	3	(4)	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	(21)
		29									30											
Hydrocarbon	22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
		31									32											
Ketal	22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	(20)	21
		33									34											
Ketone réagent	22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
		35									36											
Epoxy	22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	(17)	18	19	(20)	21
		37									38											
-O hydrocarbon	22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
		39									40											
-O acyl	22	(1)	2	(3)	4	5	6	7	8	9	10	(11)	12	13	14	15	16	17	18	19	(20)	(21)
		41									42											
-O-hetero	22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
		43									44											
-N-hetero	22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
		45									46											
S-hetero	22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
		47									48											
Miscellaneous	22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21

FIGURE 2.

60		*61*		*62*		*63*	
⓪	O-Acyl	0	O-Hetero	0	N-Hetero	0	S-Hetero
①	Carboxylic	1	Morpholine	1	Morpholine	1	Thiophene
2	Poly	2	Furan	2	Piperidine	2	Thiazole
3	Unsat	3	Lactone	3	Pyridine	3	
④	Aromatic	4	Spirostane	4	Pyrimidine	4	
⑤	Aliphatic	5	Sub in 0 spiro ring	5	Pyrrole	5	
⑥	Substi	6	Pseudosapo.	6	Thiazole	6	
⑦	St. chain	7		7		7	
8	Cyclo-alkyl	8		8		8	
9	Branched	9	Misc.	9	Misc.	9	Misc.
11	Hetero-cyclic	11		11		11	
12	Inorganic except hal						

64		*65*		*66*		*67*	
⓪	Bile Cpds.	⓪	Sterols	⓪	Hal.	0	
①	Acids	1	Ergosterol	1	Fl	①	Androstane
②	Cholanic	②	Cholesterol	②	Br	2	Addition compound
3	Norcholanic	3	Vitamin D_3	3	I	3	Maleic adducts
4	Bisnor-cholanic	4		④	C1 Double Bonds	④	Pregnane
5		5		5	5 (6)	⑤	21 Unsubstituted
6		6		6	5 (10)	6	21 diazo
7		7		7	8 (9)	7	
8		8		8	8 (14)	8	
9	Misc.	9	Misc.	9	1 (2)	9	Misc.
11		11		11	1 (10)	11	
12		12		12		12	

FIGURE 3.

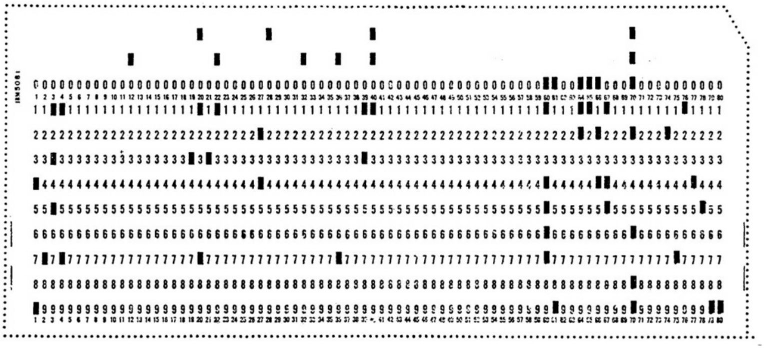


FIGURE 4.

The Matrex device

Whereas, in the IBM punched card system, each card represents an item (a compound or patent) on which descriptor codes are punched and is thus an instance of conventional code grouping; in a Matrex system, each card represents a descriptor on which are recorded the codes for compounds or patents characterized by that descriptor. That is to say, a Matrex system, like the Peek-a-boo system and many similar systems, uses inverted rather than conventional grouping of codes.

For this experiment a Matrex card having room for 10,000 compounds or patents was chosen (Fig. 5). It will be noted that the illustrated card has a numbered tab in this instance, 19 (9 in the 2nd decade). Since the indexing system requires 24 descriptors×22 positions, (528 Matrex cards) plus 60 cards for general descriptors, 588 cards are required for the complete system.

The 60 cards representing general descriptors are in one group; and the cards representing the 24 positional descriptors are separated into 24 sets of 22 cards. Each of the 60 general descriptors is assigned a number, and the 60 cards are grouped randomly. The numbered cards representing the positioned descriptors are also filed randomly within eight groups, each group consisting of three numbered sets of 22 as in Fig. 6.

The random filing of cards within groups made possible by the use of Radex tabbed cards, is not integral to all uses of the Matrex system; nor restricted to it, but in this application to positions of descriptors it serves to decrease materially the time required to select and refile cards. Items are entered into the Matrex system by selecting the appropriate descriptor cards from the file and drilling a hole in all the selected cards in the position corresponding to the number of the item.

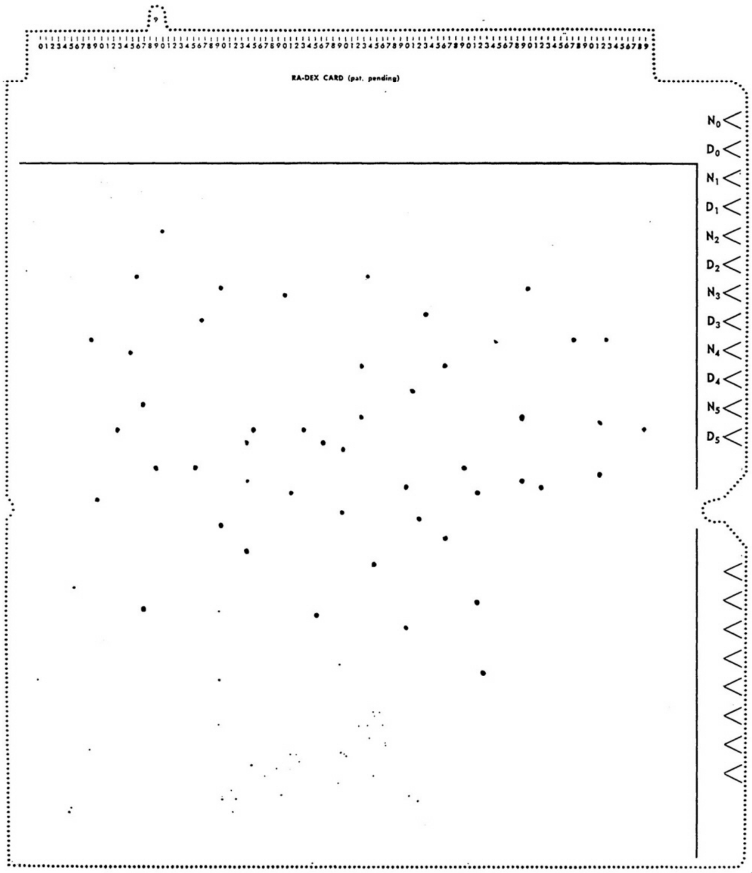


FIGURE 5.

The patent illustrated in the table (Fig. 2) and the punched card (Fig. 4), which can be considered item 1, is entered by selecting from the Matrex file=, 4, 9, 17;   1, 3, 5, 11, 17; COOH 20; etc. After all cards are selected they are placed together in the drill jig and a hole is drilled through the entire set at position one.

It is obvious that any items selected by any descriptor or combination of descriptors in the IBM system will also be delivered by the Matrex system; and that any item not delivered by the IBM system will also not be delivered by the Matrex system. A sample diagram illustrating the complete convertibility of code grouping is shown in Fig. 6 which compares a conventionally grouped Zatocoding system with an inverted group of codes having the same coding space.

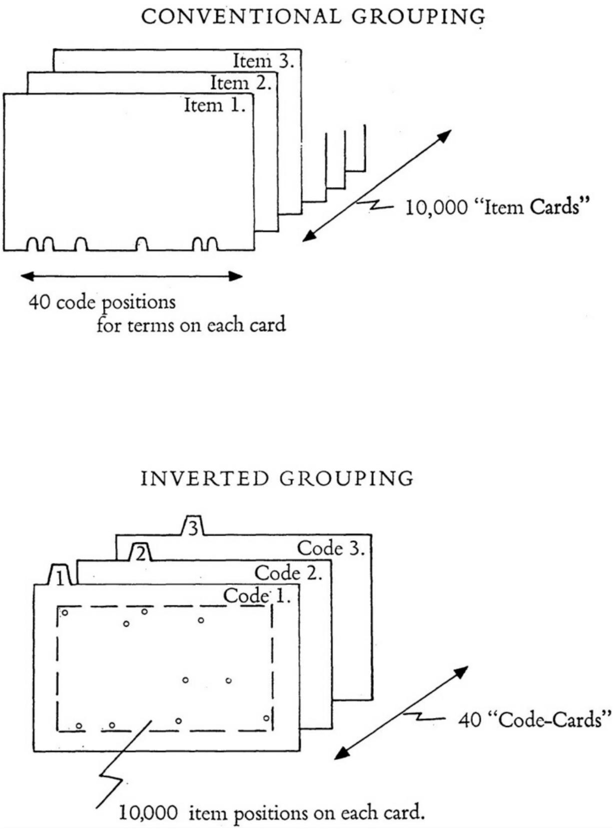


FIGURE 6.

The identity of coding between conventional grouping and inverted grouping is secured by matching the number of code positions on the Zatocard to the number of cards in the inverted system. Similarly in the Matrex system for steroids there are as many cards as there are holes assigned for use on a punched card.

In the following comparison, dollars have been used with reference to equipment; and "times" for operations for which salary rates are unknown. In all cases such times are exact only to a degree since stop watch measurements were used in some cases and not in others. However, the magnitudes are sufficiently revealing for our purposes.

Basis for comparison

In order to arrive at some basis for comparison of the Matrex and punched card apparatus for searching steroid patents, the Patent Office staff provided ten hypothetical queries and ten input problems were derived from 10 randomly selected patent cards. It might be thought that with identical indexing and identical problems, a conclusive comparison could be made. Actually, the experiment was carried out under other than ideal or perhaps even typical conditions. For example, the Patent Office group did not have readily available a tabulator which would have enormously simplified the print-out of the answers. Further, a Type 82 sorter (650 cards per minute) was used whereas a Type 83 at 1000 cards per minute would have given better speeds. On the other hand, the unique Census Bureau multicolumn sorter with dial-set wiring was used instead of the more common Type 101 which would be somewhat more costly to operate because of the more tedious plugboard wiring.

Ten input and query problems may be much too small a sample for a valid conclusion although the problems were taken entirely at random. The entire steroid patent collection (1556 patents in the machine system) may also be too small to permit reasonable extrapolation of the experimental results to larger collections. None of these inexactitudes is hopeless, however, and anyone can qualify the results to suit his requirements and draw reasonable conclusions. Perhaps the most troublesome prediction involves the tolerable size of a search answer. The search costs in either system are seriously affected by whether a search yielding, say, as many as 100 or 500 answers is acceptable, or whether the search result must be limited to a smaller number. Moreover, re-searching to vary the size of the yield may be excessively costly under certain conditions. This will be discussed later.

Input data

The following data are for system input *after* the data have been coded for the machines. Therefore normal total input costs cannot be determined from these data. Actually, the conditions here are not typical either for the punched card or Matrex device. The punched card data do not include conversion to punched card codes (a step normally not required in a Matrex device) and the Matrex file is aided by the use of random-filing tabs which reduces the time of both location and returning of the Matrex cards to the deck.

According to the Patent Office, the time required to prepare a punched card, i.e., a patent, for the system is 2 minutes (including verification) from a code sheet. This is to be compared with the data for ten input problems for the Matrex device (Table 1).

TABLE 1 Termatrex input

Patent no.	*Terms*	*Selection*	*Drill*	*Return to rack*	*Total Elapsed Time*
2742485	24	4:14	:399	1.244	6:17
2763669	19	3:10	:451	.559	4:51
2275969	16	2:12	:249	.447	3:21
2785189	24	3:21	:212	.219	4:04
2662089	19	2:41	:223	.185	3:22
2159569	23	2:45	:202	.190	3:24
2186906	19	2:58	:161	.308	3:45
2734066	23	2:50	:219	.241	3:36
2408832	16	2:16	:154	.399	3:11
1985747	9	1:30	:149	.336	2:19
Average	19.2	2:48	.242	.373	3:48
Average elapsed time	3:48				

It can be argued that the times are more nearly equal when punching instructions must be prepared. But this is beside the point for the purposes of this paper.

Output data

The Patent Office supplied ten queries for use in this experiment. The characteristics of these queries and the results are summarized below. The Patent Office supplied a blanket figure of 6 1/2 minutes time to present the answer cards to any query of the deck of 1556 patent cards on the Census Bureau's multicolumn sorter. We can assume that somewhat more time would be required on a standard IBM 101 machine.

The results produced by the runs indicated by the Patent Office were duplicated on a Type 82 IBM sorter (650 cards per minute). In order to keep the running time to the lowest possible value, a predetermination was made of the order of search terms giving the final answer in the minimum sorting time. This was determined absolutely for this experiment. In normal practice this ideal sort may not always be achieved, although a skilled searcher might do nearly as well. An important point to remember in examining the following table is that the search yield is merely a presentation of the answers and not a read-off or print-out that may be taken to the Patent files. In the case of the IBM machines, the presentation is a group of punched cards in nonsequential array. In the Matrex device, the presentation is a series of dots of light whose coordinates must be read off. We must assume that in all instances the patent numbers must be copied down or tabulated on a sheet of paper.

In the Appendix is given the detail of the sorting operation for each query which is of interest in showing how the deck was reduced on each pass. We were hopeful that such data could be generalized to write into an equation for prediction purposes, but the sample is too small. Such an equation based on several hundred random queries would be useful for cost determination purposes.

The question of read-out is important, cost-wise. With a Matrex device it takes approximately 6 seconds per item to read the coordinates and write down the number. We find that it takes about 4.5 seconds to write down a number from an interpreted punched card. A print-out from punched cards could be achieved much faster with a tabulator, but the cost of the tabulator must be reckoned. The steroid punched card deck need not be maintained in numerical order for search purposes, but the answer cards should be sorted for read-out purposes.

TABLE 2

Query	*Number of terms*	*Number of answers*	*Searching time, minutes*			*No. of cards handled in Sorter*
Matrex	*101*	*Sorter*				
1	4	363	1.5	6.5	7.5	3495
2	7	155	2.3		8.1	3434
3	7	148	2.4		8.8	3633
4	5	18	2.4		6.5	1701
5	6	61	1.8		6.2	2120
6	4	17	1.2		6.3	1690
7	5	3	1.0		5.8	1739
8	8	69	2.3		8.5	2744
9	6	344	1.4		8.3	3887
10	7	15	1.7		6.0	1737
Total			18.0		72.0	
Average	5.9	119.3	1.8	6.5	7.2	2618.0=1.68×deck of 1556 cards

All these factors must be related to the size of the answer. This would be of no concern if all searches yielded answers such as query Nos. 4, 6, 7, and 10 in Table 2. But query 1, with 363 answers presented, is another matter. With the Matrex device it would take 36.3 minutes to write out the answers; with punched cards, 27.2 minutes *after* a numerical sort. A tabulator could print out the answer from punched cards in about 6.3 minutes.

A tabulator is one of the more expensive pieces of equipment and a relatively high load factor is necessary to justify it.

If a presentation of 363 answers is too large, the remedies are: *(a)* addition of another term, *(b)* substitution of terms, *(c)* complete rephrasing of the query. The first is easy to do with all three devices but probably simplest with the Matrex device. The second is still easy on the Matrex device, and considerably more difficult on the punched card machines. Rephrasing takes us back to the relative costs indicated by the time of search in Table 2.

Equipment costs

Table 3 shows the cost of equipment required for input and searching with punched card and Matrex devices.

TABLE 3

Punched card (sorter)		*Punched card (101)*		*Matrex device*
Punch	$2,067.00	Punch	$2,067.00	Light box, template, and drill (10,000 items capacity) $285.00
Sorter	6,943.00	101	31,800.00	

If a tabulator were added to the punched card system for read-out purposes an additional cost of $29,627.00 would have to be considered.

Extrapolation to larger collection

Theoretical considerations have pointed to the unfavorable effects of growth of a collection in a system not using inverted grouping. There follows a brief analysis of this effect as exemplified by the punched card steroid index. Many intangible and policy questions enter into the problem of how many answers a search device should present for digestion by the inquirer. It is impossible to say that he should or should not go to the file and examine X number of actual documents. If the volume is large, print-out aids are necessary and the investment must be faced. By stopping short of the print-out question, the effects of collection growth can be brought sharply into form.

Let us assume that the steroid patent collection grows to 10,000 punched cards. (This happens to be the limit of a cheap Matrex device; more sophisticated Matrex devices can be had for collections of 40,000 (for between $1000 and $1500). The input costs of either punched cards or Matrex are not affected by growth, if we ignore the relatively low cost of tabulating card stock.

The story is quite different on the output side. The cost of answer presentation on the Matrex is the same whether 10 or 10,000 items are in the collection. With conventional punched cards the cost of search goes up linearly with the size of the deck. In examining just how much, we encounter the most interesting phenomenon that the more sophisticated 101 approach suffers greatest from growth and becomes far less efficient than an ordinary sorter.

If the 101 takes 6.5 minutes to search a deck of 1556 cards, it would take 41.8 minutes to search 10,000 cards.

Assuming equal proportions of drop out and the same operating efficiency as on the smaller deck, 10,000 cards would require

$$\frac{10{,}000}{1556} \times 7.2 \text{ minutes} \quad \text{or} \quad 46.3 \text{ minutes to search}$$

The operating efficiency on the smaller deck is

$$\frac{2618 \text{ cards}}{7.2 \text{ min.} \times 650 \text{ cpm}} \quad \text{or} \quad 56\%$$

It is reasonable to assume that, with 10,000 cards, the operating efficiency may be raised to a figure of 80%. Thus the search time would be equivalent to

$$\frac{.56}{.80} \times 46.3 \text{ minutes} \quad \text{or} \quad 32.4 \text{ minutes.}$$

Table 4 summarizes the times involved in entering material and searching the present collection of 1556 steroid patents and a hypothetical collection of 10,000:

TABLE 4

		Time for answer presentation per query, minutes	
	Entry time per patent, minutes	*Collection of 1556 patents*	*Collection of 10,000 patents*
Matrex device	3.5	1.8	1.8
Type 82 sorter	2	7.2	32.4
IBM 101	2	6.5	41.8

ACKNOWLEDGMENT

The authors are indebted to Mr. Don Andrews and Mr. Julius Frome of the U.S. Patent Office for their cooperation in supplying certain basic information for this paper.

REFERENCES

1. TAUBE, M., Cost as the measure of efficiency of storage and retrieval systems. *Studies in Coordinate Indexing,* Vol. III, Chapter II, pp. 18–33, 1956.
2. *Ibid.,* p. 18.
3. *Studies in Coordinate Indexing,* Vols. III and IV.
4. FROME, JULIUS, and JACOB LEIBOWITZ, A punched card system for searching steroid compounds. Patent Office Research and Development Report No. 7, July 8, 1957.
5. *Ibid.,* Fig. 2A.
6. *Ibid.,* Fig. 2B.
7. *Ibid.,* p. 3.
8. *Ibid.,* Fig. 4.

APPENDIX DETAIL OF SORTING SEQUENCES USED IN 10 SEARCHES

Query	*Column*	*Row*	*Drop out*	*Total cards handled*	*Time*
1	67	1	689		
	20	7	514		
	1	4	373		
	21	3	363 answers	3495	7:12
2	40	12	453		
	22	1	321		
	1	4	281		
	22	11	280		
	21	3	278		
	60	7	262		
	60	5	155 answers	3431	8:04
3	67	5	657		
	20	7	395		
	22	11	367		
	66	5	209		
	39	3	151		
	67	4	150		
	1	5	148 answers	3633	8:91
4	19	2	68		
	67	1	22		
	20	7	19		
	1	4	18		
	21	3	18 answers	1701	6:44
5	2	6	210		
	22	1	97		
	67	5	71		
	1	4	63		
	22	11	62		
	21	3	61 answers	2120	6:12
6	6	7	67		
	22	1	25		
	20	7	25		
	19	3	17 answers	1690	6:18

Query	*Column*	*Row*	*Drop out*	*Total cards handled*	*Time*
7	36	1	79		
	35	9	72		
	1	6	26		
	1	1	3		
	1	4	3 answers	1739	5:59
8	20	1	439		
	40	12	278		
	3	5	84		
	22	11	83		
	60	7	79		
	60	5	78		
	60	0	78		
	19	3	69 answers	2744	8:73
9	22	1	531		
	20	12	402		
	20	7	364		
	1	4	345		
	22	11	345		
	21	3	344 answers	3887	8:49
10	40	6	66		
	67	5	35		
	20	7	18		
	67	4	17		
	60	7	15		
	60	5	15		
	60	0	15 answers	1737	6:22

The Evaluation of Systems Used in Information Retrieval

CYRIL CLEVERDON

Recent years have seen a two-pronged attack to deal with the problems which have been caused by the immense growth in the amount of recorded information, the greater complexity of the subject matter and the increasing interrelationship between subjects. First, there have been many attempts to devise new indexing systems which will be an improvement on the conventional methods and, on the other hand, a great deal of work has been done in developing the mechanics which can be used, from the simpler kinds of hand-sorted punched cards to high-speed computing machines.

Several theoretical evaluations have been made of the various systems, but it appears that the position has now been reached where it is necessary to make a practical assessment of the merits and demerits of information retrieval systems. A project which will attempt to do this has been started under the direction of the author with the aid of a grant from the National Science Foundation to the Association of Special Libraries and Information Bureaux (Aslib).

Previous work undertaken by Thorne and the author (1) was useful chiefly in making apparent the main factors that had to be taken into account. It became obvious that the only practicable method of comparing various systems is on the basis of their economic efficiency. Any system can, if economic aspects be disregarded, reach a high level of retrieval efficiency, even if it involves looking at the majority of individual documents in the collection, and the important matter is to find which system will give the required level of efficiency at the lowest cost. It is useless to attempt to compare any two established indexes unless one also has reliable data concerning their compilation costs.

There are three main items to be considered in the costs of information retrieval, namely the cost of indexing, the cost of equipment used, and the cost of retrieval. The indexing cost is influenced by the salary paid to the indexer and the average time spent in indexing. Included in the cost of equipment are all the charges involved from the time when the indexer makes his decision until the stage where the record has been entered and put into the form which

CYRIL CLEVERDON The College of Aeronautics, Cranfield, England.

permits another person to make a search, whether this has been done by typing entries onto catalogue cards, punching holes in manually or machine-sorted cards, or any other method that may be used. The retrieval cost must cover not only the time costs involved in searching the index, but also the time cost involved in making physically available the required document or documents.

These three aspects are closely interrelated, and decisions taken regarding one point will affect the others. It must, however, be emphasised that not always are they entirely dependent on each other. For instance, much of the early criticism of the Uniterm system of coordinate indexing was not against the system, but was based on the difficulty of comparing long lists of numbers. Different and improved methods of recording the indexing decisions have invalidated all such criticisms, and the basic merits or demerits of the Uniterm system, as an indexing system, are the same whether the mechanics used involve a search time of 30 minutes or 30 seconds.

This project is, therefore, concerned only with ascertaining the efficiency of the systems as such. Obviously it will be necessary to use catalogue cards or some other method, but it is intended to separate this aspect. Certainly there is no intention of carrying out any comparative tests on the mechanics of information retrieval. Much work of this nature has already been done, further work is planned by other organisations, and it should be simple to build in the results of such work to the results from this project.

THE INDEXING PROGRAMME

In brief, the programme envisages the indexing of 18,000 aeronautical papers by four different systems; three indexers will be employed, and strict time controls will be maintained.

In devising the programme for the project, certain arbitrary decisions had to be taken, the first of which obviously concerned the systems to be tested. The four systems selected were: (1) the Universal Decimal Classification, (2) an alphabetical subject catalogue, based on the Special Libraries Association “Subject headings for aeronautical engineering libraries,” (3) a special faceted classification, (4) a coordinate system based on Uniterm.

Other systems were considered, but it is felt that these are reasonably representative of present practice. For instance, two systems are long established, two are recent innovations. Two are classification systems, while the other two are not. Two systems are conventionally used for collections covering all subjects, while the other two are intended more for specialist collections.

The results will be strictly valid only for the systems used, but it is thought that it should be possible at a later stage to test other systems without necessarily having to duplicate all the work put into this project. It could also be argued that the results will be valid only for the subject field used, but here again the expectation is that it will be possible to test in other subject fields with less effort. The field of aeronautics has been selected for this project, mainly because of the availability of documents in the author's library, but it has certain other attractions. Not only does it cover a broad field of knowledge, but it also has aspects of great specificity. Subjects in the broad field would include, for example, mathematics, mechanics, fluid and aerodynamics, heat, light, sound, electricity, theory of structures, chemistry and meteorology among the sciences, as well as electrical, military, mechanical and production engineering, metallurgy and fuel technology. Half the documents to be indexed will cover this broad field, while the other half will be restricted to papers dealing with high-speed aerodynamics (Mach number >0.8). This intense concentration in one detailed subject area, combined with the general coverage of a wide range of subjects should show the varying capabilities from the viewpoint of special or universal systems.

The ability of the indexer is of paramount importance, yet very little attention has been paid to the problems of actual indexing. Whereas one very vocal school of thought insists that technical indexing can be done only by a person with technical qualifications, there are others who argue that this is untrue, and that anyhow it would be wasteful to employ technical persons on such work. However, no results are known of tests designed to compare the average indexing ability of different types of persons. Potential indexers for this project might be described as being within the following broad groups: (A) technical knowledge of the subject plus indexing experience; (B) technical knowledge of the subject but no indexing experience; (C) indexing experience in the subject field; (D) theoretical knowledge of indexing; (E) neither theoretical nor practical knowledge of the subject nor of indexing.

In deciding the types of persons to recruit, it was considered that there were very few persons with the qualifications outlined in (A) and that such individuals would not be likely to join the project. The significance in certain circumstances of using an individual such as outlined in (E) was appreciated, but it was decided not to use this category in the project. Three indexers, therefore, have been recruited to be representative of groups (B), (C), (D). The first man, with several years industrial experience, has taken a 2-year postgraduate course and received his diploma for a thesis in aerodynamics. The second man has been a Fellow of the Library Association since 1952 and for four years has been Librarian of a large aircraft firm. The third man has recently passed the examinations of the Library Association in classification and cataloguing, but has no practical experience of indexing technical literature. This variation of experience will enable a number of useful comparisons to be made, and it will have its effects on the economic costs, since the salary one might expect to pay (in the United Kingdom) would be in the ratio 5:4:3.

A main policy decision that always has to be made is the depth of indexing and, other things being equal, this will affect the time spent in indexing. Few reliable figures have been given for current practices, although a particularly high figure is the 1 1/2 hours average quoted (2) for indexing reports for the catalogue of aerodynamic data prepared by the Nationaalluchtvaartlaboratorium in Holland. It appears from personal discussions that an average of 20 minutes for a general collection of technical reports is the top limit, and this has been taken as the maximum indexing time to be used in the project. The staff will always index the documents as fully as possible within the set time limits which will be, with various groups of documents, 20 minutes, 15 minutes, 10 minutes, 5 minutes, and 2 1/2 minutes.

It is necessary both to index sufficient documents so that retrieval is not too simple, and also to continue the indexing long enough to give a comparison of the indexers' rates of learning with the different systems. For financial reasons, it is unnecessary to continue an experiment beyond the stage where there is no significant change in the test results. It is believed that the indexing of 18,000 documents will meet these requirements.

As will be seen, in this one programme we are attempting to evaluate three variables, namely the system, the type of indexer, and the indexing time. With four systems, three indexers and five indexing times, the number of permutations is sixty. The documents to be indexed will be divided into groups of one hundred (from here on referred to as a “document group”), and therefore 6000 documents will be indexed before the same indexing conditions are repeated. By that time, however, all the indexers will presumably have become more adept in using the various systems.

The procedure will be that the first indexer will index documents 1–100 by system A, allowing himself an average time of 20 minutes for each document. Immediately after indexing a document by system A, he will allocate the appropriate headings or classification numbers for systems B, C, and D, but this will be done without any time control. Documents 101–200 will then be indexed by system B with the 20-minute allowance for each document, followed by the postings for systems A, C, and D. Items 201–300 and 301–400 will be similarly indexed by systems C and D.

This procedure will be repeated for documents 401–500, 501–600, 601–700, and 701–800, except that for these document groups the average indexing time will be limited to 15 minutes. For documents 801–1200, the time will be limited to 10 minutes; for 1201–1600 it will be 5 minutes, and finally for documents

1601–2000 the indexing time will be limited to 2 1/2 minutes per document. Meanwhile the second indexer is carrying out a similar procedure with documents 2001–4000, and the third indexer is doing the same with documents 4001–6000. The indexing of documents 6001–12,000 will repeat the conditions found in the indexing of 1–6000, and the whole stage will be repeated in documents 12,001–18,000.

The type of documents within each group will be kept as nearly as possible similar to those in any other group. The first requirement is that one-half of the documents should deal with the specialised subject field of high-speed aerodynamics while the remainder range over the broader subject fields. However, we think that the project may show up some interesting facts concerning the comparative "indexibility" of various documents. There are some people who believe, for instance, that American reports present more indexing problems than British reports, but that there is no difference between articles in American or British periodicals. A typical document-group, therefore, is made up as follows:

Research reports	
National Advisory Committee for Aeronautics	25 papers
U.S.A. industrial organisations	5 "
Royal Aircraft Establishment	20 "
British industrial organisations	5 "
Periodicals	
Journal of Aeronautical Sciences	11 articles
Royal Aeronautical Society Journal	7 "
Jet Propulsion	8 "
Aircraft Engineering	6 "
Aircraft Production	5 "
Communication and Electronics	2 "
Interavia	2 "
Product Engineering	2 "
Metal Progress	3 "
Quarterly of Applied Mathematics	4 "

In order to try to make some assessment of the ability of the indexers under controlled conditions, it is intended that at various stages in the project a selected list will be made of documents that have been indexed. This list will be sent to a number of organisations and individuals who have expressed their willingness to cooperate, and they will be invited to index these documents by one of the four systems. The individual or organisation concerned either will have technical knowledge of the subject field or will be specialists in the system in which they index or will combine both attributes. Since no time restriction will be placed on this indexing work, it might be reasonably assumed that the indexing should be correct (if such a thing is possible). Entries for this indexing would be made in the various catalogues, with an identification mark so that it will be possible to find in the testing whether these entries give better retrieval than the entries of the project's indexers.

THE TEST PROGRAMME

Although the main testing will not be done until the indexing work has been completed towards the end of 1959, it is hoped that some preliminary testing will be possible in time to present the results at the International Conference.

The first point to be emphasized is that when completed the four catalogues will be permanent tools which can be used as often as required and for a number of different tests. If it is felt that one series of tests fails to produce the type of result required, then it will be perfectly possible to make any further tests that can be devised. Complete testing is, however, certain to be a long process, and it is desirable that as many factors as possible should be covered in a single series of tests.

Five variables have been introduced into the indexing, namely the indexer, the system, the indexing time, the experience of the indexer, and the size of the collection. All these are considered to be of importance, and the tests must show the differences. In the testing we can introduce at least two new variables, the first of which concerns the types of questions put to the index. These may range through all grades of specificity from the (comparatively) broad question such as “safety considerations in aircraft design” or the narrower question such as “shear stresses in oblique plates,” down to the most specific type of question for which there can only be one correct answer, a question such as “hinge moments of a horizontal tail, 45° swept back plan form, of aspect ratio 2 and taper ratio 0.5.”

The second variable concerns the types of persons who are physically attempting to retrieve information, and these can be summarised as follows: (1) the originator of the enquiry, whose knowledge of the indexing system might be either reasonably good or non-existent; (2) the technical indexer; (3) the librarian indexer; (4) a technical man who did not originate the enquiry; (5) a librarian who had not been engaged on the indexing; (6) any combination of groups 1–5.

The testing is based on the procedure of putting the same questions to the different indexes, and comparing the resulting answers, but within this general statement there lies a number of possible variations. At this stage it is not possible to say the extent of the testing that will be necessary. Each document group will be a unique collection and will therefore have to be tested as such. It will be necessary to put to each document group a sufficient number of questions to obtain an average result which is reasonably valid. The size of the figure is not known for certain, but it appears unlikely to be lower than ten, and on this basis it would be necessary to ask 1800 questions to test the whole collection of 18,000 documents. Preliminary tests will be made to ascertain this accurately, before the major test programme starts, but for the purpose of further discussion in the paper, it will be presumed that ten is the figure decided on.

In theory, the questions may be either genuine or faked; that is to say, they could be either questions which were raised in the normal course of the work of an aeronautical establishment, or they could be questions devised for the purpose of testing the project indexes. The former method was used in the Astia-Uniterm (3) tests, and this method carries with it the corollary that there may or may not be in the index an answer to any particular question. The Astia-Uniterm tests were inconclusive because agreement could not be reached as to which answers were correct and which were irrelevant. Since no one has suggested any way in which this problem can be overcome, the first series of tests will be done on the principle of putting questions which are based on documents known to be in the indexes.

A sufficient number of people, not directly associated with the project, will be asked to compile questions based on documents in the collection. If, as is hoped, thirty people cooperate, each person would be asked to compile a total of 60 questions. They would have complete freedom of choice within selected document groups, but would be asked to vary their type of question over the three grades of specificity mentioned earlier.

This procedure ensures that there is always at least one right answer to each question, and it may be presumed that the system which produces most "right" answers in given circumstances will be shown to be the most efficient. This is not, however, certain, and deeper analysis than this will be necessary to discover the relative economic efficiency of the systems, since here other factors have to be taken into account. As stated earlier, the retrieval cost must include not only the time costs in searching the index but also the time costs involved in making physically available the required document. This latter item is a figure which will vary with each individual organisation and will have to be built into the results. Examples of possible answers will explain this point.

Four systems A, B, C and D when asked the same very specific question produce answers as follows:

System A produces document card 1	in 5 minutes
System B produces document cards 1, 2, 3, 4, 5	in 3 minutes

System C produces document card 2	in 5 minutes
System D produces document cards 2, 3, 4, 5, 6, 7, 8	in 4 minutes

Document card 1 refers to the document on which the question was based and is therefore the "right" answer. System A can thus be debited with a search time of $(5+x)$ minutes (x being the time taken to fetch the required document).

System B produces references to five documents, but it is not possible to say which is the document giving the required information. Therefore the documents have to be fetched and scanned. On an average, it may be presumed that the correct document will be found halfway through such a search, and therefore this system produces the right answer, but it is debited with a search time of $(3+3x+3y)$ minutes, where y is the time taken to scan a document.

System C fails in a time of $(5+x+y)$ minutes, and system D fails in a time of $(5+7x+7y)$ minutes.

In different organisations, x is a figure that will vary. When the time taken to fetch a document is one minute, the difference between system A and system B will be small, but should x be as high as five minutes, then the total search time would be only 10 minutes for system A as against a minimum of 20 minutes for system B.

The result can be considered differently when the question is of a general nature. Although there will always be the "right" answer, there possibly will also be other answers which would, in practice, be equally useful to the enquirer, and some extra allowance should be made for systems which produce these correct answers. The main difficulty in doing this is that it brings one back to the rock on which the Astia-Uniterm tests foundered, namely the personal interpretation of what is or is not a relevant answer. The only way to avoid this dilemma is for the main emphasis always to be placed on the "right" answer.

However, as Fairthorne states (4) "search for objects known to exist, and for those that may not exist, cannot be done efficiently in the same way." This is a theory that it might be possible to test by putting to the indexes a sample of genuine questions.

ANALYSIS OF TESTS

With so many variables being introduced into the indexing and the retrieving, there will be many differing answers given by the results, and it is unlikely, to say the least, that one system will be preeminent irrespective of the conditions. The test results will be a set of objective statements showing what has been achieved under the differing conditions, and it will be for each organisation to consider the results in the light of its own requirements and select those conditions which are the most favourable.

The first basic question that has to be decided is the minimum acceptable retrieval efficiency, and it is often too readily assumed that this must approach 100% as closely as possible. To strive for a figure such as this would appear to be reasonable only if the collection indexed also approaches 100% of the available pertinent information on the subject. With very few exceptions, such as patent offices, which, by the nature of their work, will have a complete file of a certain type of recorded material, the large majority of organisations concerned with science or technology do not have in their collections more than a small fraction of the potentially available information in all aspects of the organisation's work. Within strictly limited subject fields, it is possible that a librarian might manage to collect 50% of the potentially available documents, but it will require extensive and prolonged searching to raise the figures over 80%.

All special librarians are continually faced with the problem of marginal information; it arises in the acquisition of books, periodicals, or reports and is equally difficult in the decision as to whether an in-coming document should be indexed, ignored, or passed to the waste-paper basket. Such decisions must in the first place be based on a broad policy decision, even though there will be differing individual interpretations. One major organisation bases its policy on the externals of the documents it receives, discarding, for instance, all interim reports, preprints, or reprints or anything written in a foreign language. As a result it eliminates all but 35,000 documents of the 250,000 documents received annually. More often the decision is based on subject content, and while in certain fields published abstract journals will be relied on, there will inevitably be exceptions. An aeronautical organisation might decide to ignore, for indexing purposes, papers dealing with materials, and rely instead on *Metals Review, Crerar Metal Abstracts, Titanium Abstract Bulletin,* etc. Exceptions would have to be made for report series which, perhaps for security reasons, could not be included in published abstracts, and also it might be considered desirable to include in the organisational index all articles dealing with a subject of particular local importance, such as fatigue of certain metals. Therefore over many subject fields of possible interest to an organisation there has to be the economic decision as to whether indexing is justified. Too often the argument is put that failure to find information locally might result in a major cost to the organisation. The argument should really be between the cost of finding it with an internal index and the cost of finding it by other means.

Another local factor that can influence the interpretation of the test results is the frequency of use of the catalogue and its relationship to the number of documents indexed. While I know of no figures which purport to show what the relationship should be or even what it actually is in any particular case, in my own library it appears to be in the ratio of 1:5. This is roughly calculated, comparing the known number of items annually indexed (8000) with the estimated total of successful uses of the catalogue. By sample checks it has been shown that about 1300 annual successful uses of the catalogue result in the issue of approximately 2100 reports. A number of these issues will duplicate previous successful uses of the catalogue, and this number is believed to be about 500. This leaves 1600 items used of the 8000 actually indexed. The clear implication of this is the necessity to debit each use of the catalogue with the indexing costs for five documents, and illustrates the importance of the relation of indexing time and retrieval time. If the ratio is 1 to 5, as in the example, it would pay to decrease the indexing time by five minutes a document, so long as retrieval time was not increased by more than twenty-five minutes.

Yet another factor to be considered is the average potential value of information which is retrieved or, alternatively, the average potential cost of failure to find information.

It will be quite practical to work out and publish hypothetical examples covering these various aspects, but in the end each organisation will have to make decisions based on its own circumstances.

THE MECHANICS ASPECT

As emphasized earlier, the project is concerned with the system rather than the mechanics which might be used, and it was stated that the costs of the latter could readily be built into the results. This decision was taken partly because it would greatly increase the project cost to prepare different type indexes, but also because of the work already done and other investigations now in progress, in particular at the Forschungsinstitut für Rationalisation in Aachen. We were, however, faced with the problem of deciding on the mechanics to be used for the various systems, and how they can be adapted to yield further information.

For the U.D.C. catalogue, the alphabetical subject catalogue and the faceted classification catalogue we shall use conventional 5 by 3-inch cards. Most organisations issuing aeronautical reports include with each document, catalogue cards showing the subject headings or classification numbers for the report. An analysis of these cards shows that on average about 4 placings are given to each report, while less than 2% are given more than eight placings. Basically, we wish to test the two conventional systems (U.D.C. and the alphabetical subject catalogue) in the way in which they are conventionally used, and therefore it would be reasonable to confine our entries to the general average given above. Such a restriction, however, might well invalidate some of our investigations into the most economic indexing time, yet to give the indexers carte blanche authority to call for a very large number of entries might again upset the economic calculations. No practical investigations appear to have been done which support this conventional figure of about 4 entries for an aeronautical report, and presumably the practice has become as it is because of a general feeling by indexers that such a number of entries is adequate, and that further entries would not compensate for the extra costs involved. We may be able to find whether there is any justification for these beliefs. For this purpose, the indexers will be instructed to enter on the master cards all subject headings or class numbers which they consider in any way suitable, but they will place a mark against those which are considered to be essential and which would be allocated in normal practice. Such entries will be identified in the card catalogues, and in the testing it will be possible to find whether the additional entries are justified by reason of improved retrieval, or whether they clutter up the catalogue and are positively harmful by increasing the number of irrelevant answers.

The coordinate system should present no problems in this connection, nor should the faceted classification, although in the latter case there will be the necessity of compiling a chain index.

CONCLUSION

The project is an attempt to make a practical contribution to solving some of the problems of information retrieval. The major difficulty in evolving the programme has been that at present absolutely nothing can be taken for granted; there is no single fact which can be demonstrably shown to be true; no theory put forward by one expert which is not refuted by another. Experience in other fields of knowledge shows that most worth while scientific achievements result from long series of tests and rarely from a single experiment, and that repeated testing is necessary to determine reliably the nature and relative effects of all relevant variables. The experiment in which the researcher knows for certain that all the variables have been identified and whether each will significantly affect the results is the exception rather than the rule.

It is most unlikely that this project is “the exception.” We have, however, included in the programme at least seven controlled variables, so that the tests may fairly be described as a “series of tests,” and we may reasonably hope that the results will materially advance our present knowledge of what is and is not important in information retrieval. We also believe that even when all possible testing has been done, the catalogues and indexes will remain of value as a yardstick by comparison with which the testing of other systems and in other subject fields will be a comparatively simple process.

REFERENCES

1. C.W.CLEVERDON and R.G.THORNE, A Brief Experiment with the Uniterm System of Co-ordinate Indexing for the Cataloguing of Structural Data. (R.A.E. Library Memo 7), 1954.
2. K.STAPLES and M.SHAW, The N.L.L. Card Catalogue of Aerodynamic Measurements. (R.A.E. Library Memo 17), 1954.
3. D.E.GRAY and A.R.C.ASSOCIATES, Report on Reference Tests of Conventional and Uniterm Systems. ASTIA Reference Center, Library of Congress, 1954.
4. R.A.FAIRTHORNE, Information theory and clerical systems. *Journal of Documentation, 9,* 113 (1953).

Experience in Developing Information Retrieval Systems on Large Electronic Computers

ASCHER OPLER and NORMA BAIRD

THESIS

Large commercially available electronic computers are readily adaptable for handling a very wide range of information retrieval problems. We define information retrieval as an operation performed on a stored file of individual items containing coded or classified descriptions of their referrent. The operation consists of selecting those items which satisfy a given set of search criteria and then presenting the individual references to the searcher. In the electronic computer systems that we have been developing, the file has consisted of magnetic computer tape on which are stored coded individual items. The retrieval process consists of a sequence of operations under electronic computer control designed to select automatically every item which meets all search criteria. As an adjunct to the search process, the reference portion of the retrieved item is machine edited and printed in some convenient form. Our experiences have shown that the use of electronic computers provides fast, convenient, accurate, and inexpensive retrieval.

There is a very wide range of information retrieval systems, and it would be naive to claim that electronic computers would be best for every case. We have concluded that the three determining factors are the size of the information file, the frequency with which questions are posed for retrieval, and the complexity of the criteria by means of which desired information is selected. One may have a small collection of information which is infrequently consulted and whose indexing is extremely simple. Here, the simplest use of an alphabetized notebook would suffice. As the scope of the system becomes larger, unpunched, edge-notched and machine-sorted cards play intermediary

ASCHER OPLER and NORMA BAIRD Computer Usage Company, New York.

roles. When the system has become very large and requires frequent reference to logically complex comparisons, the electronic data processing machine is best suited. The boundaries between the regions where various levels of mechanization should be employed are rather vague and will depend not only on these three basic factors but also on the resources available to the organization.

One apparent anomaly is the role of the small and intermediate sized computer in this area. Experience has shown thus far that little gain is experienced in going from primitive methods such as machine-sorting to computers of this size. While the reasons are many, the two most obvious ones relate to the input speed and the speed of executing a single command. Most of these machines use perforated paper tape or punched cards which have inherent reading time slower than most sorting machines. Assuming that the information content of a document can be stored in coded form on a single punched card (a serious limitation), the maximum input speed will be two hundred documents per minute. At such a rate, one could equal this with five successive passes through a modern sorting machine. The best paper tape readers generally available will scan information at approximately the same rate as the 200-card-per-minute reader. A paper tape input device commonly used on several low-cost computers will read at the rate of only six medium-length-records per minute.

In complicated cases, the input speed may not be the true limiting factor. In programming the retrieval process, the manipulations and decision-making processes require hundreds and often thousands of commands (many of them repetitive) to be executed before a document can be rejected or accepted. In one retrieval problem actually run, up to two thousand steps were required by such a machine, and a decision rate of only twenty-five documents per minute was achieved. In general, the small machines carry out instructions in milliseconds whereas the large machines operate in microseconds. It is this manyfold difference in computing speed, as well as the vastly superior input speeds (e.g., 9,000 documents per minute), that make the large computer practical for information retrieval and the use of smaller computers marginal.

CRITERIA FOR MECHANIZED INFORMATION RETRIEVAL SYSTEMS

Before describing our experiments, we would first like to consider what one should expect from a satisfactory information retrieval system. Since the viewpoint of those involved will be different, depending on their interests, we present the criteria in this fashion. These are stated at this time to provide the reader with a reference background for evaluating systems to be described.

FROM A USER'S VIEWPOINT

The man who will use a mechanized information retrieval system by asking questions and receiving reports will be primarily interested in three things. He will want the results of the search to be completely accurate, including all pertinent information and excluding all else. He would like to see the results with all possible speed. He should not be required to disturb his normal routine by studying new techniques, or traveling considerable distances to pose his questions. The user would like somehow to reserve the privilege of browsing through the file if it were possible, since one function of a collection of documents is to serve as a stimulant toward new approaches to his problems.

FROM A DOCUMENTALIST'S VIEWPOINT

For simplicity, we classify the documentalist as the person having the responsibility of selecting a system (coding scheme, retrieval logic, the machine itself, and lines of communication to and from the machine) and implementing it in daily practice. This will include the design of the system and the computer program at the outset, and the subsequent routine maintenance and searching of the file. He would like to be as free as possible in developing the retrieval system that best meets the needs of his organization. (We expect that any new system he creates will be assisted by the machine rather than limited by it. With such systems, it should no longer be necessary to conform to standard classification schemes, but instead, the freest rein should be given to imaginative concepts.)

Certain entirely subsidiary but practical considerations are also desirable. Once the machine retrieval project has been started, a considerable amount of daily activity will be required. The design of the system must be such that these can be performed with the utmost of convenience. Therefore, the manner in which the searching and file maintenance entries are prepared must be as simple and direct as possible. Since space is often at a premium in areas where records are being consulted, it is important that no unduly large, bulky, noisy equipment be installed on the site.

FROM AN ADMINISTRATOR'S VIEWPOINT

The introduction of a mechanized retrieval system into an organization will pose new problems for the administrator, whether or not the organization maintains a computing installation. If the organization does not operate a computer, he will have to enter into contractual negotiations to rent computer time elsewhere. More serious problems arise when an organization currently operates a data processing machine. The first consideration will be to see that the retrieval function does not seriously interfere with the performance of the “normal” computing tasks. As an administrator, he will also be interested in keeping the actual cost of the retrieval work to a minimum. He will want to be satisfied that the value obtained is commensurate with the cost of the operation.

REPORT ON EXPERIMENTAL WORK

The limitations of punched-card systems for realistic retrieval problems led us to start testing the potentiality of electronic machines for this purpose. In the past two years, we have formulated, programmed, and tested several information retrieval systems for two different large machines. The following is a description of three major projects undertaken.

A. In contrast to many proposals for comprehensive, universal classifications, Robert G.Heitz of the Dow Chemical Company's Western Division Research Laboratory proposed an alternative scheme consisting of a large array of small, specialized individual classifications where no pattern of conformity need be established. A proposal was later made to provide for automatic inter-conversion of classification schemes in areas of overlapping interest.

It is at once obvious that only a machine with a complex logical ability, in addition to a large memory and a flexible input facility, can cope with this problem. In concept, one might have the file of each specialist stored on a separate reel of tape, preceded by a computer program tailored to retrieve from the file according to the specialist's own classification.

As a case study, the classification schemes and the files of two chemists, A.A. and F.N., were programmed. A.A. used a random-number superimposed coding scheme. He had prepared a lexicon of approximately two hundred terms and assigned ascending sequences of four two-digit random numbers to each term. His card file, which bore serial numbers, was originally edge-punched with the codes of all applicable descriptive terms. His method of retrieval was to sort out from the deck those cards containing all the punches corresponding to the combined codes for all the words in a search question.

The other man, F.N., made the fullest imaginable use of the potentialities of edge-notched cards. Within the framework of ninety-six permissible punching locations, he included direct coding, indirect coding, numerical coding, alphabetical coding, and matrix coding. His retrieval method used needle sorting to reduce the card deck to a small number of cards that could be hand-inspected. This reduction was accomplished by using whatever combination of techniques he deemed necessary.

The original concept of a reel of tape for each man's program and file was set aside for this case study. Instead, both files, converted to a new format, were stored on a single reel of tape with a suitable separation mark. A single program was devised for the IBM 702 Electronic Data Processing Machine to search both files, retrieving according to the corresponding request form and delivering a printed, edited output for each requester. Some of the details of the computer programs might be of general interest and are outlined below.

1. In no case was it necessary to alter or modify the logical selection method used by the specialist. For each man, a punched-card format was devised which embodied all the elements of his search question. This card, when fed to the computer, caused the search program to examine each record in the file for logical correspondence.
2. To increase searching efficiency, provision was made to search the file with as many as five independent requests from each of the two men. This mode of operation, termed multiplexing, makes more efficient use of available equipment.
3. With the equipment used and the multiplexing arrangement, search rates of 2,500 (A.A.) and 1,000 (F.N.) comparisons per minute were attained. Each of these comparisons matched one independent request with one record and required from six to eight major logical decisions before acceptance or rejection of the record. For information purposes, the basic rate of logical operation for this machine is 0.14 millisecond. When operating at maximum search rate, approximately 25 milliseconds are required for the complete logical comparison of a record. Since an average of one hundred computer commands must be executed to reach a decision, it is clear that we are utilizing electronic speeds properly, although with improved programming and searching techniques, the rates cited above could be increased many-fold.
4. For the random-number superimposed codes, the conversion to magnetic tape records was extremely simple. Each tape record consisted of the identifying serial number followed by the digits corresponding to the original holes (e.g., 17643–11–17–29–, etc.).

   The conversion of the edge-notched card file proved to be a challenging and arduous task. This required the development of a separate computer program for this purpose alone. To the computer were fed punched cards containing the serial number of the record followed by a list of the sequence numbers of the holes that were punched. The computer program interpreted and digested the information and produced tape records suitable for subsequent searching. (A separate account of this conversion procedure is to be prepared for publication.)
5. The retrieved records were collected internally, sorted so that separate

pages for each of the ten searches could be prepared, edited to produce a printed line containing that information desired by the requester, and then written directly on magnetic tape. By the use of tape output, computer operation was not slowed down. In our project, we made use of an IBM 720 device which prepared our printed output directly from magnetic tape at a speed of five hundred full-width lines per minute.

B. A project to mechanize the retrieval of stored chemical structures has been reported elsewhere (1–5). We wish to refer to this project in this context to the extent that it may be considered a good example of an information retrieval project carried out on a modern computing machine.

In the system as presently constituted for the IBM 704 Electronic Data Processing Machine, a large file of coded chemical structures stored on magnetic tape is searched at high speed, and the names of compounds complying with the search request are retrieved and printed. Among special features of the system are the retrieval of partially indeterminate structures, alternative classes, and the use of multiplexing. Search speeds in the order often thousand structure comparisons a minute are obtained.

Since the last published description, two variations of the output program have been developed. In one, the names corresponding to the retrieved serial numbers are found by a special tape search, and these are edited so that the output for each multiplexed question appears on separate pages. Another option under development involves the search for special codes which depict the structure on a cathode-ray screen and photograph it. Thus, the results of a search for chemical structures are portrayed on a film strip returned to the requester (6).

C. In contrast to the two preceding projects, which were undertaken to retrieve actual information, our third project was conceived as a fundamental study of the theory of information retrieval. We sought (1) to illustrate the relationships between symbolic logic and retrieval and (2) to demonstrate efficient computer techniques for manipulating both conjunctive and disjunctive phrases.

The "file" of information to be searched consisted of a collection of six-letter words. These words were made up of random letters (e.g., MVALBB). Each word could be considered an expression in symbolic form of the conjunction of these six classes. The search requests consisted of symbolic logical statements about letters and, when correct retrieval was obtained, the six-letter word itself was printed out. The order within a word had no significance.

While the input and output formats were quite simple, the logical statement in the search request could be extremely complex. Permissible items in the question were elements, disjunctive classes and conjunctive sub-classes within classes. These could occur in any combination in both normal and negated form. Thus, with this combination of possibilities, we were able to pose any logical request that we could conceive. An example of a permissible request is:

> a word containing no vowel; one of the first ten letters of the alphabet; no letter consisting entirely of straight lines, unless it be K or V; Q must be present and P must not.

An example of a word meeting this requirement is QVCBQS.

This retrieval scheme was programmed for the IBM 704. Since it was essentially an experimental program, no multiplexing was employed. Approximately 1,200 six-letter words were stored within the computer. Searching rates varied with the complexity of the questions and ranged from 15,000 to 18,000 words per minute, exclusive of input-output time.

EVALUATION OF ELECTRONIC COMPUTER SYSTEMS IN THE LIGHT OF CRITERIA

We now wish to review the criteria for satisfactory information retrieval systems in relation to the achievements and shortcomings of the three systems described in the preceding section.

FROM A USER'S VIEWPOINT

Criterion: Accuracy

As our technology advances, we are confronted with ideas of ever increasing complexity. To describe such concepts requires the freest use of logical associations. To restrict retrieval to elementary combinations of logical conjunctions will not usually suffice to define the desired concepts and, thus, the retrieval system able to cope with the greatest logical complexity should be expected to have the greatest logical accuracy. We see no reason why it is necessary to accept the principle that some uncertainty will always be present. The chemical search problem illustrates the freest combination of retrieval methods, including simple correspondence, the use of numerical ranges, logical manipulations, and the difficult problem of topological connectivity. Our complex logic program demonstrates how the most realistic logical combinations can be built up to obtain as fine a discrimination as desired. Once the retrieval system has been selected, the accuracy of commercial machines surpasses that of any other information retrieval equipment publicly described.

Criterion: Speed

The speed which concerns the user is the total elapsed time between asking his question and getting his answer. In addition to the actual machine running time, serious delays can occur unless the arrival of the question coincides with the availability of the machine. Proper scheduling to minimize delays is the function of the computer administrative group and involves techniques beyond the scope of this paper. To summarize, the total elapsed time, at best, should be less than one hour for a moderately complex system involving up to 100,000 records. At worst, poor synchronization in handling may delay the results up to twenty-four hours.

Criterion: Convenience

Electronic systems are certainly convenient from a user's viewpoint. With any modern machine, the output format can be arranged into the most intelligible array of numbers, symbols and English words. Should these prove insufficient, the output may be arranged in a pictorial form, as mentioned above.

In our work, we have relied upon a simple translation procedure for preparing the punched cards which activate the searching program. This has not proved difficult, but should it be deemed worth the effort, an automatic translating scheme could be programmed.

Criterion: Browsing

The subject of browsing with a mechanized information system has not, to our knowledge, been investigated. While the creative process is, of course, a human function, machines can be of considerable assistance. We visualize the implementation of two forms of automatic aids to browsing. The first relates to the accidental discovery of an ostensibly unrelated item that is suggestive to the user. In this, the machine is directed to a broad area and instructed to draw at random a given number of sample records. After perusal by the user, he may call for the drawing of additional samples in the same area or, if he has found an interesting clue, he may direct that the area of interest be accordingly narrowed and further samples drawn.

In the second, a method that relates to controlled association may be programmed. In this, a single area of interest that has been exhaustively covered is fed to the machine. The computer program locates and analyzes statistically the concepts or descriptions associated with the original area. The machine then samples from the directly associated areas and, if so instructed, from areas indirectly connected by two or more direct associative links to the original.

There is little doubt that these procedures can be programmed as adjuncts to information retrieval systems. This area seems a profitable one for future study. This might serve to answer critics who claim that mechanizing information retrieval promotes intellectual sterility.

FROM A DOCUMENTALIST'S VIEWPOINT

Criterion: Versatility

The introduction of a mechanized retrieval system, while providing enormous help to the scientist, will prove a challenge to the alert documentalist The transition will require the learning of some new techniques, a re-evaluation of existing methods, several basic policy decisions and considerable arduous and detailed work. Those organizations that now have a large collection of documents classified according to some satisfactory scheme will probably want to mechanize the system unchanged. Those with new or unclassified collections must first select the most suitable mechanized system for their application. Fortunately, electronic computer systems can cope with *any* rational system selected. Our Logic Program demonstrates this conclusively.

Several cases demonstrating the ease of modifying classifications according to some unambiguous principle have arisen in our work. In one case, a single category was divided into two on the basis of the identity of associated items. In another case, the meaning of an inclusive class definition was replaced with a somewhat different one. Furthermore, after a program is written, the fundamental logical retrieval scheme may be altered or augmented at will, as is always the case with computer programming.

Criterion: Convenience

Once the format of the information file has been selected, the meticulous task of originating the basic machine file must be performed. In practically no case will this be easy. The material must first be read, digested, indexed, and entered onto magnetic tape directly or via intermediate punched cards. Work is now under development to make some of these processes automatic. This original magnetic tape must now be processed by the computer, using an editing program, to produce the working tape with the desired format.

After the original collection has become part of an operating system, a procedure must be instituted for adding new documents to the file periodically. This is a standard operation for installations working with large files, and the techniques have been well worked out.

In the general and in the chemical retrieval systems described above, we wrote several file editing and file maintenance programs. These proved relatively simple to write.

Criterion: Compactness

Many people fear the physical intrusion of large pieces of machinery into the library. We wish to disavow the notion that this is necessary. In the section below, "From an Administrator's Viewpoint," we will discuss some practical problems and demonstrate that remote operation is completely feasible. Under these circumstances, all that is necessary in the library is a remote inquiry station. This may be merely a telephone, an intercommunication system, or a keyboard device which transmits coded questions to the machine area. As an example, we propounded five chemical search questions in California, telephoned these to New York where the searches were run (six minutes) and received the results within three hours.

Since the magnetic tape files will not be stored in the library, their bulk should be of no direct concern there. It is of interest, however, to note the extreme condensation attained. With the magnetic tape and the tape units that we have used, we can place 30,000 to 90,000 documents on a single reel, where these range from 215 to 60 characters respectively. With the new magnetic tapes and new tape handling units now appearing on the market, an increased condensation of at least two and one-half times is expected. Thus, the volume required to store the indices to one million documents will be slightly more than one-fourth cubic foot.

FROM AN ADMINISTRATOR'S VIEWPOINT

Criterion: Non-interference

Since it is a practical necessity that a large machine be shared (see the section below on cost), it will be important to schedule the actual machine retrieval to conform to the overall machine use pattern. Where the daily operation consists of a series of short computations, it is simple to intersperse retrieval operations with these. Where long runs are standard, it will usually be necessary to wait for the conclusion of the day's processing before the searches can be run. Where the relative importance of rapid information retrieval is realized, it is feasible to prepare the program for the long runs to allow for controlled interruption. In this manner, urgent demands for information could be satisfied and then the normal program resumed.

While most of our experience has been gained on data processing machines assigned solely to rental users, we have also made considerable use of a machine engaged in daily processing using long runs. In this case, we were particularly fortunate since these long runs were carried out during the evening shift, thus freeing the computer for short runs during the day.

Criterion: Cost

The cost of installing a large electronic data processing machine is extremely high. The machines capable of carrying out searches at high speeds will rent from $20,000 to $50,000 per month, or may be purchased for more than a million dollars. This is not as alarming as it sounds when one considers the enormous flexibility and work capacity of such a machine. In addition to information retrieval, these machines can do literally hundreds of different types of operations, ranging from the highly mathematical to the simple commercial. In only a few highly specialized cases could the installation of a large computer solely for searching be justified. These exceptional situations arise where very large files must be frequently searched (e.g., U.S. Patent Office, *Chemical Abstracts,* etc.)

Since exclusive control of the machine is not generally economical, it is obvious that one must usually share the machine. This sharing is relatively simple for those affiliated with organizations having their own computers. The large machines will usually be encountered in business organizations having very large financial accounting problems or highly technological problems. In addition, many large universities and government organizations have sufficient use to acquire large computers. In general, all such organizations will have some available time, and its use by information groups is quite in order.

For those groups with significant information retrieval problems, but whose management does not plan to acquire a large machine, it is quite practical to rent facilities by the minute. Such data processors might be available at computing centers specially established for the rental business or at other organizations willing to sell machine time. Remote operation is widely used and the communication problem has been generally solved.

In the case of sharing one's own machine, the pro rata cost will approximate $2 to $5 per minute, while direct rental will average $5 to $12 per minute. The fact that searches seldom will involve over ten minutes should be kept in mind. All the researches described under “Report on Experimental Work” were done by renting equipment for $5 to $11 per minute.

The initial operating cost of information retrieval by large computers is not low, but the cost per search may be. There is a relatively large expense that must be incurred before the first search can take place. These expenses include the cost of preparation of the information files for the machine, the transcription to magnetic tape, the writing of the searching program and the necessary supporting programs, the correction and editing of the data files, and the location and correction of programming errors. In general, this initial preparation will depend on the length of the file and the complexity of the retrieval program. In no case will this be less than several thousand dollars.

Assuming machine costs of approximately $5 to $12 per minute and searching comparisons at the rate of 2,000 to 20,000 per minute, one can predict a base cost of roughly 200 to 4,000 comparisons per dollar. To this base cost must be added the proportionate cost of the initial phases. Where the number of searches is expected to run into the thousands, this distribution will add little to the cost per search.

In addition to the originating expenses and searching expenses, the cost of adding new information must be included. Here again, the cost per new entry-will consist of a low fixed cost per addition plus a proportionate share of the expenses in preparing the updating program.

Thus, it appears that the operating costs of machine retrieval will depend on the length of the file, its rate of growth, its complexity, and the frequency of searching.

CONCLUSION

The commercially produced electronic computer in the information retrieval field has yet to be accepted. Only the experience to be gained in the years ahead can fully justify its widespread adoption. Our work has explored several areas and several techniques that should be practical. Many of the problems that appear insurmountable turn out to be relatively simple when study and imagination are brought to bear on the machine. It is hoped that studies in this field will be intensified in the years to come.

BIBLIOGRAPHY

1. A.OPLER and T.R.NORTON. New speed to structural searches, *Chemical and Engineering News,* vol. 34, pp. 2812–2816, June 4, 1956.
2. A.OPLER and T.R.NORTON. *A Manual for Coding Organic Compounds for Use with a Mechanized Searching System,* The Dow Chemical Company, Midland, Michigan, 1956.
3. A.OPLER and T.R.NORTON. *A Manual for Programming Computers for Use with a Mechanized System for Searching Organic Compounds,* The Dow Chemical Company, Pittsburg, California, 1956.
4. A.OPLER. A topological application of computing machines, AIEE Special Publication T-85, *Proceedings of the Western Joint Computer Conference,* 1956.
5. A.OPLER. Dow refines structural searching, *Chemical and Engineering News,* vol. 35, pp. 92–96, August 19, 1957.
6. A.OPLER and N.BAIRD. “Display of Chemical Structural Formulas as Digital Computer Output,” from a talk before the Division of Chemical Literature, 133rd National American Chemical Society Meeting, April 1958.

Printing Chemical Structures Electronically: Encoded Compounds Searched Generically with IBM-702

W.H.WALDO and M.DE BACKER

OBJECTIVE

In a paper entitled "Routine Report Writing by Computer" by Waldo, Gordon, and Porter published recently in *American Documentation,* announcement was made of an industrial research report writing system where a large-scale computer functioned as its core. The paper made reference to a new method of encoding the structures of chemical compounds for automatic storage and retrieval. The input from the raw data was simple enough to be handled by clerks. The output was a printed structure recognizable and acceptable by a trained chemist. Furthermore, the computer was programmed to permit generic searches to be made.

This paper describes the details of the structure coding, methods of searching, and retrieving, as well as a discussion of some of the things that we have been able to do and plan to do in the future. It is hoped that this paper will be the basis for discussion, further research by others, and whatever changes that are necessary to bring about a full-dress review of the whole field of chemical structure codes and ciphers. The chemical profession is sorely in need of a generally accepted method for indexing the structures of organic chemicals.

DISCUSSION OF OTHER SYSTEMS

In 1945, Dr. D.E.H. Frear of Pennsylvania State University (1) generated a code which through the cooperation of many competent chemists was adapted by the National Research Council's Chemical-Biological Coordination Center. This represented the first full-scale attack in this country on the problems

W.H.WALDO and M.DE BACKER Monsanto Chemical Company, St. Louis, Missouri.

of correlating chemical and biological data and searching for the members of a family of compounds encoded from the literature.

Since that time many ciphers have appeared which were designed to permit the indexing and searching of chemical compounds, such as those of Dyson, Wiswesser, Silk, Gruber, Wiselogle, and Gordon-Kendall-Davison.

To evaluate these systems the cooperation of one hundred chemists throughout the United States was secured to encode and decode compounds by several of these systems (2). The results of the experiment were particularly revealing in that not only were the ciphers not chemically unique but it was especially difficult to regenerate the original compound from the cipher.

Since the appearance of these ciphers, two other systems of recording chemical structures have been announced that are based upon a topological principle. Topology is involved also in the system described in this paper, which, incidentally, was independently conceived prior to the announcement of either of the other systems.

These two systems that involve the principles of topology are those of A.Opler and T.R.Norton (3) and L.C.Ray and R.A.Kirsch (4). Both the systems are in operation and further attention is directed to them for comparison with the system described in this paper.

The principle of topology involved here assumes that a chemical represented by its structure consists of hard cores (the chemical elements involved) joined by rather flexible links (the several types of chemical bonds). It is further assumed that these hard cores and links flexibly connected may be twisted into a large but finite number of shapes. Long alkyl chains can be stretched or bent, hexagonal and other ring systems may be squeezed into rectangles; rings may also be stretched to permit the inclusion of bridge-heads. The various moieties may be twisted about to permit the writing of any structure that can be displayed on paper.

Our system has two features not found to our knowledge in any other. The encoding is simple enough to be done by trained clerks. The output may be a two-dimensional printed structure easily recognizable by trained chemists. Both of these advantages were specified requirements when the project was undertaken.

MONSANTO'S APPROACH

The project was undertaken at Monsanto for reasons similar to the several others previously mentioned. We are in dire need of some method to index structures and permit their retrieval by classes. The program was undertaken, strangely enough, not by a librarian, nor by a document custodian, but by a technical editor. The reason lay in the belief that the solution to the critical structure problem would open up the entire field of chemical documentation to machine handling.

Every discussion of machine documentation in the field of chemistry ultimately results in a discussion of chemical codes and means of recognizing selected groups of chemicals from the rapidly increasing list of known compounds. Three years ago when the project was in the discussion stage, we at Monsanto felt that once the structure storage problem appeared to be solved, the solution should be only a part of an entire machine documentation system.

We also felt that our research program should proceed in small increments, each increment being reduced to practice before investing in the next increment. Such a procedure, a typical industrial approach, guaranteed a maximum return for a minimum of investment.

There was another design criterion in our philosophy of approach. This criterion involved the belief that any mechanized documentation system was doomed to economic failure if it was not a total system. Partial systems have been studied for many years, and the results of their work have laid the basis for this conclusion. The system must be designed around the capacity of the machine rather than try to force a machine to do some segment of the classical technical information system.

Reports written for immediate consumption are difficult to index for posterity because of their local jargon. Papers written in Japanese or Russian are difficult to read by Americans unless the readers are especially familiar with the foreign language. Thus total communication is not effected by journal publication. It is, furthermore, generally conceded that no individual scientist is interested in the mass of chemical literature, making critical abstracting virtually impossible. What may be absurd and formless to one man may be the key to the enigma in the mind of another. Also, authors of chemical articles are especially adept in contributing knowledge in an incidental way, incidental, that is, to their primary purpose in writing, so that information may be lost in the communication system, author-referee-editor-abstractor-indexer-searcher.

To avoid these blackouts in our technical communication system, we believe it is imperative that any new system that may be invented has no hope for success unless it begins with the bench worker, passes through all the necessary steps, and concludes with a future user of the recorded facts.

Thus to invent a code for chemical compounds without having in mind at the time the chemist at the laboratory bench doing research, the report writer, the documentalist, the current report reader, and the ultimate retrospective searcher is a mistake. Their interests are only in the information recorded rather than a complex code by which the information was stored, and the code never can be depended upon to produce a functional, economic, and acceptable tool for furthering chemical research. It is chemical research that we are interested in furthering. Chemical documentation, at best, is a service to chemists.

In our opinion one of the most serious drawbacks of the various ciphers and codes proposed for the storage of chemical structures is the ultimate dependence upon a book of rules that is too thick to be memorized easily. We believe that our system is easily learned and can be readily incorporated into the education and language of chemists.

No matter how appealing, how erudite, how clever, astute, or how novel a documentation system may be, if it is not acceptable or easily sold to the practicing chemists involved, it is doomed to failure. The system must be better than the present system, not in our opinion, but in the opinion of the scientists who are to be its ultimate benefactors. To satisfy management it must have some inherent factor of savings. To satisfy the authors, it must be an improvement on what they are now doing in terms of time and effort. To satisfy searchers, it must be quicker, more complete, and more accurate.

REWRITING STRUCTURES

Although our philosophy of approach is independent of the hardware, the details of this work are reproducible only with the 702 or 705. One other important factor in Monsanto's approach to machine documentation of chemical information is possibly a limitation in so far as this paper is concerned. We had available to us a large-scale electronic data processing machine, the IBM-EDPM-702. All the work has been performed with the 702 or lesser IBM equipment. We hope that our work can be adapted to computers manufactured by other concerns. Some of the work might also be adapted to medium-sized computers. However, many of the rules and conventions that may appear at first glance to be arbitrary were invented to be compatible with the 702.

We believe that this paper describes an entire chemical documentation system from the generation of the information to its ultimate use at a later date. Of course, the key is the storage and retrieval of chemical structures in a manner that meets these criteria.

It may sound impertinent to say that structures to be encoded for machine storage must be accurate, but there is no slander intended. Rather it is recognition that specialists in one branch of chemistry or another often take much for granted when they record a structure of a compound that they have just synthesized. Aromatic rings must have the double bonds shown. Empirical aliphatic chains leave the geometric isomerism in doubt. The presence of double bonds must be indicated, not assumed. For instance, —CSSR (a moiety) may represent a polysulfide or a dithioate. Since these factors are important for any search, they must be properly encoded and not left in doubt. However, our system does not require the presence of every hydrogen atom to be shown. Indeed, the hydrogen atom has no place in our storage system, and, as you shall see later, the computer has been instructed to determine the proper number of hydrogen atoms for each structure stored.

A structure is stored by first writing the structure clearly and accurately in the conventional form.

The encoding consists simply of rewriting the structure on common cross-hatched paper with numerals used for chemical bonds and single character letters for the elements. Table 1 is a list of the bonds involved.

TABLE 1

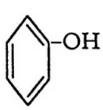
1=single bond
2=double bond
3=triple bond
7=ionic bond, designating salts, complexes, etc.
8=the point of attachment bond, where there is an indeterminate number of C atoms in a chain
9=the point of attachment is in doubt where there exists indeterminate geometrical isomerism
/=a special bond, symbol used in rare cases to indicate single bonds only.

Table 2 is a list of the single character letters that we have used to represent the elements involved in the structures. We hasten to point out that our problem is primarily with organic chemicals.

TABLE 2

C=carbon
N=nitrogen
O=oxygen
S=sulfur
P=phosphorus
X=chlorine
Y=bromine
I=iodine
B=boron
L=silicon
F=fluorine
R=any wholly covalent moiety
M=all other elements

The elements and bonds are placed in the squares of the cross-hatched paper to compare exactly with the original chemist-written structure. To show the general principle, consider the structure of phenol.



Encoded it becomes

C	1	C	2	C	1	O
2				1		
C	1	C	2	C		

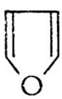
You will note the following conventions: (1) the aromatic resonance is frozen into the alternating sequence of double and single bonds; (2) there are no hydrogens in the system, thus a single bonded oxygen represents a hydroxyl group; (3) the hexagonal benzene ring, so common to organic chemistry, becomes a rectangle. This form of the organic structure is now ready for a keypunch operator. To make it convenient for her the structure must be "boxed" and punctuated, viz.

		C	1	C	2	C	1	O	,			
		2				1	,					
		C	1	C	2	C	.					

The "box" is simply the left-hand, top, and bottom limits of the structure. The punctuation is the right-hand limit. Boxing and punctuating make it a simple matter for the keypunch operator to enter the structure onto a punch-card.

All structures are handled in an analogous manner. The rules presented so far are sufficient to handle the great majority of compounds, including many steroids. However, there is a relatively large number of compounds that require some special handling.

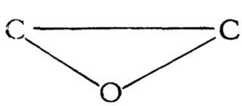
Compounds with an odd number of elements in the ring (five or above) are handled by repeating the appearance of one of the bonds three times, so that the rectangular configuration is maintained. For example:



becomes

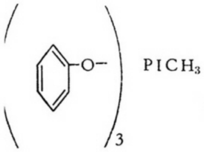
C	1	1	1	C
2				2
C	1	O	1	C

Three-membered rings and penta-substituted elements such as phosphorus and nitrogen, a relatively infrequent occurrence, are handled with a slash.



becomes

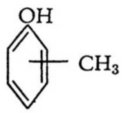
C	1	C
1	/	
O		



becomes

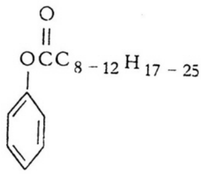
C	1	C	2	C				C		I		C	2	C	1	C
2				1				1	/			1				2
C	1	C	2	C	1	O	1	P	1	O	1	C	2	C	1	C
								1								
		C	1	C	2	C	1	O								
		2				1										
		C	1	C	2	C										

An alkylated phenol, where the position of the side chain on the ring may be unknown, is handled with the 9. For example:



C	1	C	2	C	1	O
2				1		
C	1	C	2	C	9	C

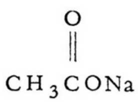
If the side chain is made from, say, a mixture of fatty acids, the 8-bond is used, and the lowest number of carbon atoms possible in the mixture is written out. Thus:



becomes

C	1	C	2	C	1	O	1	C	8	C	1	C	1	C	1	C	1	C	1	C	1	C	1	C
2				1				2																
C	1	C	2	C				O																

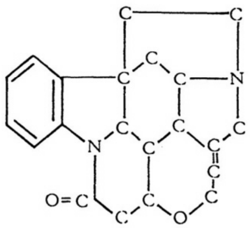
Sodium acetate shows the use of the ionic bond 7.



becomes

		O				
		2				
C	1	C	1	O	7	M

Figure 1 shows how the structure for strychnine can be handled.



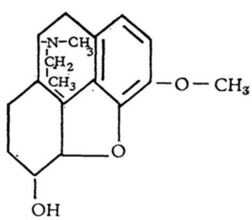
								C	1	1	1	C		
								1				1		
C	2	C	1	C	1	1	1	C	1	C		1		
1				2				1		1		1		
C	2	C	1	C	1	N	1	C		C	1	N	1	C
						1		1		1				1
				O	2	C		C	1	C	1	1	1	C
						1		1						2
						C	1	C	1	O	1	C	1	C

FIGURE 1. Encoded structure for strychnine.

Terpenes and other compounds with bridgehead structures create what appears to be a problem in geometry. But the problem was resolved by using the same principle as was used for 5-membered rings, namely repeating the bonds in the outermost rings a sufficient number of times to permit the inclusion of the bridgehead within.

Figure 2 presents the structure of codeine as an illustration. Arbitrarily we have limited the number of characters in the horizontal line to 98 since the computer is limited naturally in the length of line it can print. Furthermore, we have limited the number of rows arbitrarily to thirteen to accommodate our report-writing program (5).

To make the search program somewhat easier we have ruled further that



C	1	1	1	1	1	C	1	C				
1						2		2				
C	1	N	1	C		2		C				
1		1				2		1				
1		C				2		1				
1		1				2		1				
1		C				2		1				
1		1				2		1				
C	1	C	1	1	1	C		1				
1		1				1		1				
C		C	1	O	1	C	2	C	1	O	1	C
2		1										
C	1	C	1	O								

FIGURE 2. Encoded structure for codeine.

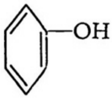
long saturated straight chains may not be bent after the second carbon.

Coordinate covalent bonds are ignored as such and are represented as a 2.

There is a myriad of ways of writing a given structure. It makes no difference how this is done so long as it is accurate. The computor will seek out all variations.

STORING STRUCTURES

In our system each compound has a number assigned. The key punch operator punches the information from the coded sheet in a continuous sequence onto 80-column IBM cards. The structure for phenol is in the conventional form:



in encoded form

C	2	C	1	C	1	O	
1				2			
C	2	C	1	C			

and keypunched in an IBM card:

column No.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	—	80	
1	2	3	4	5	1	C	2	C	1	C	1	O	,	1				1	,	C	2	C	1	C	.			

code

The first five columns of the card are punched with the compound identification number. Column 6 is a sequence control position for card numbering, and the balance of the card is for structure data storage. A maximum of ten cards may be used for each record, allowing storage of 740 characters of structure data.

The program or instructions for the computer are stored permanently in a small deck of IBM cards. All data cards to be added to the structure file are entered in the computer along with this program load deck.

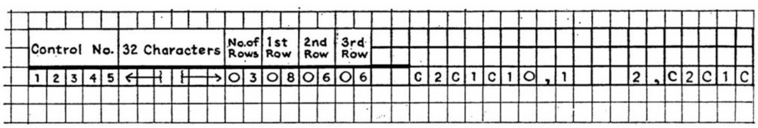
Each group of cards comprising a complete structure is transcribed through the computer to magnetic tape into the form of a variable length record. Since the record size may vary from 45 to over 700 characters, the variable length feature allows maximum reading and writing speed in computer operations.

Only significant data are sent to and from input-output devices. The maximum number of records may be stored on a reel of magnetic tape for the same reason.

As the record is being transformed from the punched card format to the output format, the 702 calculates certain control information it needs in order to remember how to reproduce the structure when called upon. It counts the number of horizontal rows in the structure (three for phenol); it also counts the number of columns in each row of the structure beginning at the left edge as indicated by the box line and terminating with the punctuation. Editing is then performed by the 702 to ensure that the structure does not contain an incorrect number of rows and that each row does not contain an improper number of characters.

Incorrectness occurs, of course, as a result of human errors. We have noted two empirical coincidences; both of which are used in this editing process. Because of the ever present sequence of elements and bonds and the sound chemical principle that a bond must lie between two elements, the number of rows in our rewritten structures must of necessity always be odd, 1, 3, 5, · · · 13. Similarly the number of characters in any row must be odd, since all elements are single-character symbols as are the bonds. However, the last position in every row is signaled with a punctuation mark. Thus the number of characters in a row is always even, two through ninety-eight. When errors are detected, the structure is rejected for entrance to the permanent file. If correct, the control numbers are stored as a part of the structure record.

Thus, for phenol, the record may be represented:



The 32 blanks are reserved for the molecular formula. The computer, after the structure is stored in this manner, proceeds to calculate the molecular formula and insert it in its proper place in the record.

As the formula is being constructed, further checking by the machine is performed to determine the accuracy of the input data. The rules established for coding structures are integrated in the program so that the computer is able to take a fairly sophisticated look at the chemist's coding and the keypunch operator's work. It will not allow any atom to have too many or too few bonds, nor is a "7" bond code permissible with atoms for which ionic bonds are not "legal." Improper atom and bond codes and misplaced characters are recognized by the computer, as are various other types of errors.

There are, of course, many errors the computer cannot possibly detect. If the chemist intended to code benzene:

C	1	C	2	C	,
2				1	,
C	1	C	2	C	.

and coded cyclohexane instead

C	1	C	1	C	,
1				1	,
C	1	C	1	C	.

the computer is blind to the error.

For the structure determined to be erroneous by the machine, a record is written showing the structure as coded, together with a message indicating the reason for rejection. These are checked by a coder and necessary corrections are made prior to re-entry to the computer.

Hydrogen atoms are not coded as a part of the structure in our system. This poses another problem for the computer in calculating the molecular formula. It must recognize that in phenol, coded:

	(A)		(B)		(C)		(D)			
	C	1	C	2	C	1	O	,		
	2				1	,				
	C	1	C	2	C	.				
	(E)		(F)		(G)					

one hydrogen atom is located at each position A through G. The computer recognizes this, and adds the necessary number of hydrogens to the molecular formula. For carbon this is easy. The infallible rule of 4 covalent bonds permits this computation without the computer being too sophisticated. The halogens also give no problem since organic halogens are bonded with only one covalent bond. Likewise oxygen has only two covalent bonds. However, nitrogen, phosphorus, and sulfur are not so easily handled, and a set of rather sophisticated rules is written into the molecular formula program so that the computer can recognize amines from azides, sulfites from sulfonates, and phosphites from phosphonates, etc.

When the formula has been calculated, the complete structure record, including the molecular formula, is written from the central computer to magnetic tape. A complete tape record has the following format:

Compound No.	Molecular formula	Controls	Structure data

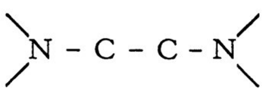
The apparently correct structures are also written in the same manner as those in error, but with no error message. These may then be scanned by chemists for errors unrecognizable to the 702. All the new structures, in an updating run, are thus printed in the original coded form. (These output structures, incidentally, are written on magnetic tape initially, and printed later with peripheral equipment, thereby releasing the computer for other work.)

One reel of magnetic tape produced by the 702 may hold over 25,000 typical structure-molecular formula records. Thus, a single file, consisting of only a few reels, contains all information necessary for searching and printing many thousands of chemical structures.

PREPARING THE SEARCH QUESTION

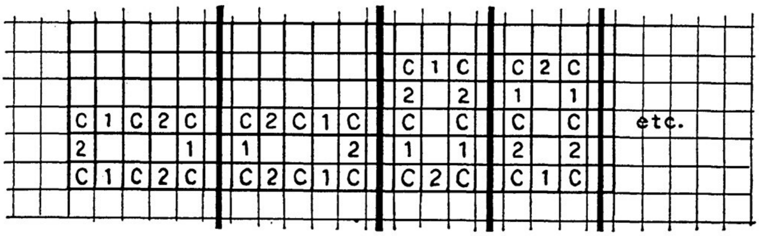
A chemist proposing a search of the chemical structure files for a specific substructure or chemical moiety must state precisely the elements, bonds, connections, and so forth, he desires to locate. The search specifications supplied are transformed to IBM cards, and become the set of rules by which the computor will perform the search.

Search questions such as finding all the derivatives of diphenylamine are meaningless to the computer. The chemist posing the question must determine for the computer what is meant by a derivative. It can pull out for the chemist all diamines, all compounds having two carbon atoms in a sequence (substituted ethylene), or it can pull out all compounds having the sequence:



Control data include the molecular formula requirements and the substructure, coded in the same manner as the structure record. Further control information is utilized to control the switching network during the search.

One coded form of the substructure is entered to the computer. As pointed out before, there are many different ways in which a group of elements and connectors may be coded. The coded substructure may appear in the structure in the form as coded on the control card. In most cases, however, it will be in any one of the many other mutations possible. A benzene ring, for example, may appear:



The difficulty in performing searches of this type of record becomes obvious. With several chemists acting in different ways at different times, similar structures may be coded in many different ways. Little education is needed for a human being to code the chemical structures for input to the computer, but this is counterbalanced by the fact that much education is necessary for the computer to locate substructures in the coded form. The advantage is, of course, that the chemist's time is available for more important duties that computers cannot perform; and our educated computer must be put to work.

The instructions for the computer can, to a certain extent, be compared with the instructions one would give a clerk who knows nothing about chemistry to scan a file for substructures. He would be told specific things must appear in certain forms, and no other type of thing should be considered. It is much easier however, to program a human being to recognize equal mirror-type images, modulo 90° turns, long bridge bonds, and such, and how to reject much of the file at a single glance. Unfortunately in this case, the computer cannot react as a human being.

OUTLINE OF THE SEARCH PROGRAM

The program deck, containing all the instructions necessary to perform any type of search, is entered into the computer with control cards necessary to perform the specific search.

We have available for high-speed scanning the molecular formula file, which is part of the structure record. As each structure is considered, the molecular formula is compared with the "molecular formula" of the substructure or moiety sought. If the substructure contains strange elements, the structure being compared is rejected by the machine. Furthermore, if the search requirements demand a greater number of any given element than is presented in the particular stored structure, it will be rejected by the computer.

If the desired molecular values are available in structure, then the detailed search begins. With the aid of the control cards entered with the search program, the computer gets its ideas as to whether certain mutations of the substructure desired should be considered. It is a waste of time for the computer to re-search the structures for a 180° turn if this is identical to the input substructure. If the substructure desired is a ring of carbons connected by single bonds, such as

C	1	C	1	C
1				1
C	1	C	1	C

the mirror image of the structure, as well as the second 90° turn, is identical with the original form. It will be necessary to search for the first 90° turn, since it takes on the form

C	1	C
1		1
C		C
1		1
C	1	C

The third 90° turn need not be tested as it is identical with the first.

Depending on the nature of the search question, and control information given the computer, the search continues until a match is found, or until the whole structure has been scanned, no equality found, and the structure rejected.

Figure 3 is the basic flow of the manner in which the 702 attempts to find what it wants.

RETRIEVING

In addition to the structures, the names and certain laboratory test results of many thousands of compounds have been stored in our system. The method of retrieval must be sufficiently variable to make optimum use of this large mass of data and its versatile availability. The data are stored in three different "files," name file, structure file, and results file. Physically these files are reels of tapes. Retrieving this information for use is, of course, dependent upon the search question. However, there is a variety of output techniques available, not only because of the versatile 702 hardware, but also, because of the flexibility built into the manner in which these files have been stored. The versatile 702 permits output in the form of punch cards, tape, or paper printed by a high-speed printer. The printer can give a single copy, multiple carbon copies, or paper mats for offset printing. The data may appear in any of these forms

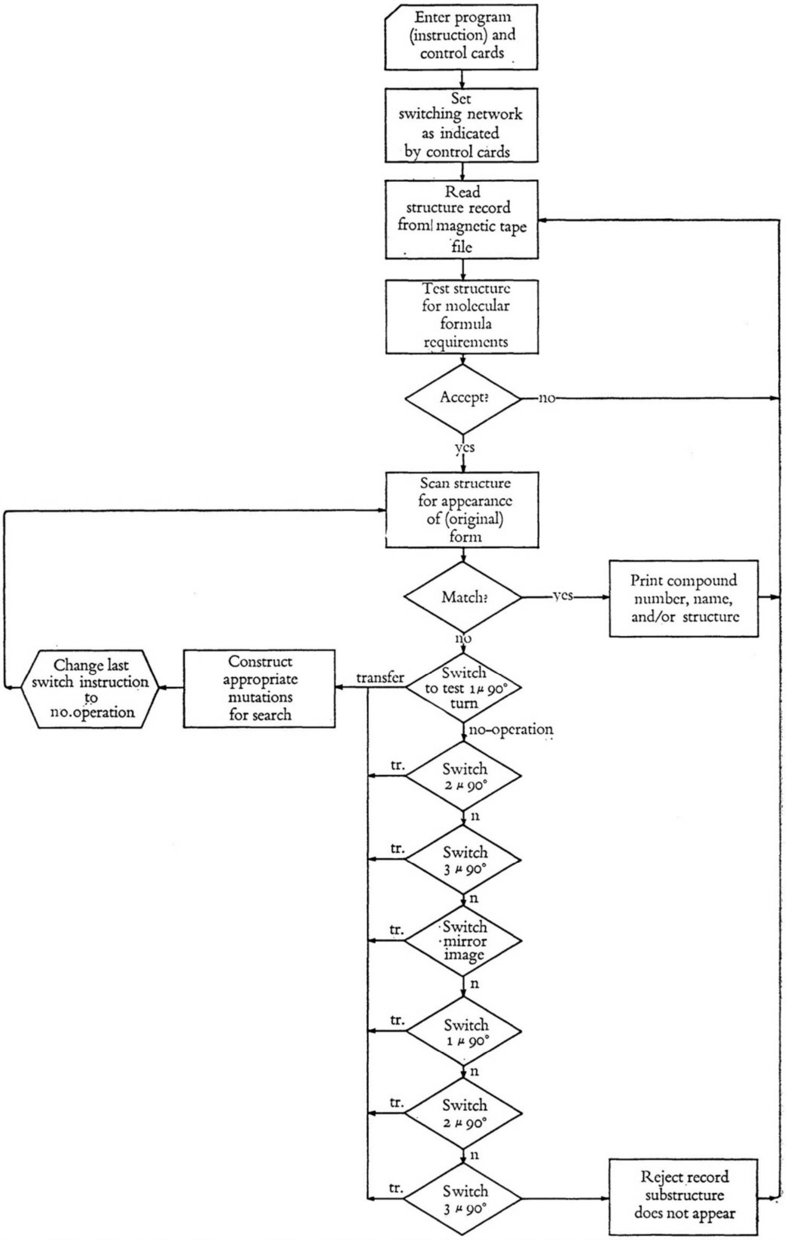


FIGURE 3. Flow diagram. Searching and printing chemical structures.

from merely a list of compound identification numbers to relatively sophisticated research reports.

The computer uses the compound serial number as its primary identification tag. Once the computer is aware of the serial number of the compound it is seeking, it can then be instructed to print out either a *Chemical Abstract's* systematic name, its structure, or any or all of the test results stored.

Because of the limited number of different characters available on the printer, certain conventions had to be adopted for the complex organic nomenclature involved. Thus all Greek letters, *o, p, m, s, n,* etc., had to be written out in their English equivalents as: alpha, ortho, etc. Similarly subscripts and superscripts appeared as, super-2, and sub-4. Parentheses, also unavailable, appear in our names as #.

Some names of complex compounds are too long to be printed on one line, so that some ingenious programming was necessary to hyphenate these names.

Retrieving the structures in a printed form is a relatively obvious step. Since the computer knows the number of rows in each structure, the number of characters in each row, and a punctuation mark signalling the end of each row, the computer is simply instructed to pick up each row and print it in the proper sequence on the page. In this manner the structure is regenerated identically with the input. It has taken surprisingly little practice for chemists to be able to identify these computer-printed structures with the same ease as any chemist does with the classical structures. Our Monsanto chemists have assured us that they have found the computer-printed structures very easy to interpret, and they feel that the price of changing the conventions from the classical to the computer-printed form is a small price to pay for the ability to make generic searches among thousands of compounds and to have the resulting structures immediately available for study.

In accord with our stated policy at Monsanto of making this research work in documentation pay for itself in each step along the way, we have been successful in using the computer to write one-page laboratory reports from data supplied to the keypunch operators from the laboratory notebooks of several scientists evaluating uses for compounds as they are synthesized.

Printing these reports requires the name file, structure file, and test results file to be fed to the computer in accord with the appropriate program. The output is a single sheet of paper showing the serial number of the compound being studied, its name, structure, all the raw data recorded, and a brief evaluation. Further details of the use of our 702 for routine report writing are given in the paper “Routine Report Writing by Computer,” by W.H.Waldo, R.S.Gordon, and J.D.Porter (S).

PREPARING FINAL REPORT TABLES

After building a considerable file of the names and structures of organic chemicals, spending over a year storing application-test data for computer use, and with the experience of writing the simplest forms of reports of application testing, we found ourselves in a position to make further use of these data.

In the writing of classical reports, data are frequently recopied for progress reports, memoranda, interim reports, special reports, process reports, and final reports. To be able to record these data once and employ a machine to do the recopying is music to the ears of an industrial chemist. From the documentalist's standpoint such an idea suggests the Utopia where several reports by cooperating research groups confirm rather than contradict because all the data come from the same set of punched cards accurately reproduced by a machine. However, probably the most important advantage in making repetitive use of test data stored on punched cards is the potential of releasing thousands of man-hours of highly skilled scientific personnel from the drudgery part of report writing, that of copying reams of data over again.

We have called the requests to regenerate stored data for the second, third, or more times, “bonus questions.” It is in this area that savings from machine documentation really pay off.

We have conducted an ever increasing number of searches in answer to these “bonus questions.” Twice so far we have searched the same set of molecular formulas for remarkably similar requests from two different departments of the company for two entirely unrelated reasons.

The first search of this type, because of our inexperience, took about twenty minutes of computer time; the second less than ten. The whole operation from receipt of request to mailing of the reply took less than half a day.

The intriguing part of the molecular-formula file is that it may be searched for any combination of elements in any order. It has been suggested that if *Chemical Abstracts* were to store their molecular-formula indices in this way, they would be able to print all possible variations in book form by offset duplication at a price that all of us would gladly subscribe to or could easily afford. Such a book would permit many generic searches to be made by hand through clever processes of elimination. For example, if one searched the file for all compounds having two or more nitrogen atoms he would get all the diazo compounds in one handy stack. It is true, of course, that azides, diamines, triamines, etc., would also appear. But the file to be hand-searched would require amazingly less time than to go through the *Chemical Abstracts Index* in its present form for polynitrogen compounds.

The most frequent bonus question, at this stage in our work, has come from the testing laboratories. They have asked for their test results to be searched by electronic equipment to produce lists for them, just lists; they have asked for lists of numbers and names of compounds that were tested under conditions *B,* but not *C;* lists of compounds that showed activity of some nature *A* without having been analyzed in test *D;* lists of names of all the compounds for the past five years that have had *X* and *Y* test results in common. The requests are just beginning to come in as the men are becoming aware of how this electronic equipment can save them days and even weeks of plowing through old records to dig up information that belongs in reports. These men are especially thankful because it saves them time, drudgery, and, most important, the chance of human error is drastically reduced.

A program is now being prepared and tested to produce an even more valuable set of tables. This program will answer a further bonus question. The question was to examine all the data for a given test during 1957 and perform a minor calculation on every piece of data recorded in this test and then prepare a table of compound number and name with these calculated results on an offset master properly paginated so that the sheets, when duplicated, can simply be punched and inserted into the scientist's final report.

Of course we are not replacing the thinking mind of the scientist with an electronic and mechanical machine. It will be a long time, if ever, before any scientist is relieved of the responsibility for defining his research program and listing his objectives in the introduction of his reports. And if it ever becomes possible for a piece of hardware to replace the scientist in drawing conclusions from the results of experimentation, science will indeed have lost its luster. The computer simply prepares the tables of data in novel and interesting fashion so that the scientist has more time to be creative, analytical, and to think.

By the time this paper is being discussed at the International Conference on Scientific Information some searches will have been made for the most difficult and far reaching bonus question we have asked of our system. We will be able to say how successful we were in making generic searches from a file of organic structures.

The question at present is not whether it is possible for us to perform such a search, but rather how long did it take us, what statistics have we found, and how has it helped Monsanto's research effort.

LIMITATIONS

This paper is not describing a perfect system of chemical documentation by machine. In the first place our files are small compared with the files of *Chemical*

Abstracts. We have something under twenty thousand compounds stored, a mere sample of the compounds in *Beilstein*. To say that the problem of scale-up to encompass the world's knowledge of organic chemistry is simply a matter of dollars is wishful thinking to the realist.

Although we have been able to "teach" the computer a great many facts about chemistry, both of structure and of application, we have made no attempt as yet to teach it chemical reactions or to teach it chemical properties, although we are certain that these things can be done.

Probably the greatest limitation of all in this work is that of personnel. We would all like to make use of this and other systems. If only someone would collect all the data in science and store it properly for us to use, but two things are lacking. There are few people with the proper training who wish to spend their lifetime feeding information to a computer as a mother does her child.

Furthermore, there is the problem of economics. It will take a huge investment to store all of chemistry, but the most glamorous and profitable rewards will not be available until the storage job is up to date.

A final limitation to such a system is a subtle one. There is almost no redundancy in our system. The chances of finding erroneous data stored by accident through the human error of recording or punching or through the very infrequent machine error is entirely left to chance. We have found these errors by tedious proofreading of the output, but this can only be done on a small scale. We have found an occasional error by chance, where someone recognizes that a piece of output data is not as he recalls it should be. How serious this limitation is, only time can tell.

CONCLUSIONS AND RECOMMENDATIONS

This paper has described a total documentation system from the generation of the data at the laboratory bench to retrospective searching. The system is barely in operation, and there is much work yet to be done. Our work, so far, has convinced us that our design premises were desirable, practicable, and attainable. A chemist makes a compound, writes its structure and gives it a name. Information about this compound is processed by clerks for storage in the large-scale computer. Application scientists test this compound and they record their results in their notebook. These results are processed by clerks and stored for the computer. Individual, one-page reports are prepared for distribution automatically by the machine.

Periodic summaries of the stored data are made routinely. Special searches are prepared on demand within a few hours time. These searches may include the selection of any class of compounds definable by structural means. Rearrangement or further calculations of the data stored and the automatic comparison of millions of "bits" of information for correlation and leads for further scientific research can be made simply. Finally, generic searches of our file of chemical structures can be made and the proper structures regenerated and printed for further use by our chemists for study or report writing. We seem to be on the threshold of new techniques in the conduct of research and development work. The national problem of the shortage of scientists has given rise to the invention of this and other documentation systems designed to relieve the scientists from many of the chores that historically were the price we had to pay for the privilege of studying nature.

The job is not done. In the tradition of science it is the intention of the authors of this paper through its presentation to invite criticism, comments, comparisons, and improvements during or after this conference.

There may be other applications of the topological method we have employed here for the storage of chemical structures. It seems to us to be applicable to any series of simple systematic diagrams, such as electric circuits, or the diagrams used in architecture, anatomy, botany, and geology.

ACKNOWLEDGMENT

The authors wish to express their appreciation to a number of the members of the Monsanto organization who have contributed materially to the current program. J.D.Porter and R.S.Gordon did much of the early programming and planning. Many of the problems associated with such a radical change in documentation were solved by William Massey, Leon Cooper, and Jill Joyce. The cooperation of the technical staff of the Monsanto Research Departments in Dayton, Ohio, Nitro, West Virginia, and St. Louis, Missouri, is also gratefully acknowledged.

REFERENCES

1. FREAR, D.E.H., *Chem. Eng. News, 23,* 2077 (1945).
2. FREAR, D.E.H., *Chem. Eng. News, 33,* 2838 (1955).
3. OPLER, A., and NORTON, T.R., *Chem. Eng. News, 34,* 2812 (1956).
4. RAY, L.C., and KIRSCH, R.A., *Science, 126,* 814 (1957).
5. WALDO, W.H., GORDON, R.S., and PORTER, J.D., *Am. Document., 9* [1], 28–31 (1958).

Evolution of Document Control in a Materials Deterioration Information Center

CARL J.WESSEL and WALTER M.BEJUKI

ABSTRACT. This paper will present a detailed account of the nature of the literature control problem and the historic development of manual systems as applied in the area of deterioration prevention.

From the experiences of the Prevention of Deterioration Center, an analysis of the principles and problems in literature control will be drawn. Material for the analysis of this manual storage and retrieval system stems from an organization having an applied and technical library of approximately 30,000 documents.

Decisions involved in going from a card catalog system, to an edge-punched card, to a Batten-type card encompass comparative costs, speed, personnel requirements, efficiency of retrieval, capacity for future growth, and similar problems. These, plus decisions on the size and type of dictionary of terms to be associated with the art and technology of deterioration prevention, shall serve to illustrate the evolutionary pathway traversed.

A model Batten system is being constructed for a hitherto unindexed satellite collection of 2500 specifications in this field. It is expected that the knowledge gained from the model will assist in perfecting a system for the larger central collection.

The significance of special concepts, such as data ranking, in a general theory of literature storage and retrieval will be considered.

BACKGROUND

The Prevention of Deterioration Center is a nonprofit, scientific organization maintained jointly by the three United States armed services, by means of an Office of Naval Research contract, under the operating supervision of the National Academy of Sciences-National Research Council. The principal

CARL J.WESSEL and WALTER M.BEJUKI Prevention of Deterioration Center, National Academy of Sciences-National Research Council, Washington, D.C.

The Prevention of Deterioration Center operates with the support of the Army, the Navy, and the Air Force, under contract between the National Academy of Sciences and the Office of Naval Research.

purpose of the Center is to collect and organize in one depository the latest research information on workable technics and scientific studies relating to deterioration prevention, and to serve in a consulting and advisory capacity on such matters for the military agencies, for other government groups, and for authorized persons not directly associated with the federal government.

In the years since its inception in 1945, the Center has accumulated a comprehensive knowledge and facility not readily available elsewhere on this highly specialized subject. It is the purpose of this paper to show the evolutionary steps traversed by the literature control system of this Center during 14 years of development under periodic contract renewal circumstances. Initially a modest conventional indexing system, it moved to a more elaborate manual procedure involving edge-punched cards, direct coding, information classification and the sundry accoutre associated with such a technic. The present state of evolution is a studied effort to modernize information storage and retrieval, particularly from the viewpoint of increasing the comprehension and ability to collate information. This requires the evaluation of the requirements of the Center.

Since the literature of deterioration is voluminous, the Center maintains a library staff to collect, catalog, and index pertinent technical data, and a publications group to summarize and condense incoming technical reports, of which some 30,000 have been accumulated to date.

The functions of the PDC library are not unlike those of any library serving a consulting group in the technical field, although the subject matter is unusual, and characterized by a marked heterogeneity. The professional staff regularly scans approximately 150 journals pertinent to the field, including official society and trade journals, other subject matter journals, and a large variety of abstract services. These perusals generate in turn the need to examine, as the occasion requires, contributory publications. All reasonable efforts are made to process government reports on the subject of materials deterioration and its prevention, including those which originate as reports of contract work performed. Bibliographical rules are interpreted to fit the special needs of the publishing activities of the Center without sacrificing what we consider to be essential library standards. Through the central library of the National Academy of Sciences, the PDC library has access to books and journals belonging to the numerous government collections in Washington.

The study of our information handling methods is well circumscribed in time. The art of deterioration prevention began as a discrete study area, in terms of the Center and the United States, when it was recognized by the armed forces in the early years of World War II that equipment with definite predictable life durations, when used in temperate climates, became inoperable in a matter of weeks when shipped to the tropic or polar regions. The Navy and Army, after some preliminary work on evaluating the extent and possible significance of the problem, established a Joint Army-Navy Deterioration Steering Committee under the National Defense Research Committee. Together with the necessary subcommittees, the organization set out to analyze the needs of the armed forces and to determine the means of satisfying those needs. Subsequent to the establishment of the Office of Scientific Research and Development, the Tropical Deterioration Information Center was instituted in 1943, using the services and facilities of the George Washington University. Acting initially in a trouble shooting capacity, the Joint Army-Navy Deterioration Steering Committee and the Tropical Deterioration Information Center proved by the end of World War II that their activities had a continuing usefulness. In 1945, therefore, the Office of Research and Invention (ORI) of the Navy Department recommended a more permanent organization. With the two existent groups, the wartime Joint Army-Navy Deterioration Steering Committee and the Tropical Deterioration Information Center as a basis, the latter was reorganized. The University group became the Prevention of Deterioration Center. This activity was governed by an ORI contract within the administration system of the National Academy of Sciences-National Research Council. Subsequently the Army and later the Air Force added financial support to the ORI contract, supported initially only by the Navy, with ORI continuing to act as the contracting office.

HISTORY OF THE STORAGE AND RETRIEVAL PROBLEM

In addition to being of historical interest, the foregoing information contributes to the evolution of the philosophy of the literature storage and retrieval problem. The development of the art of information storage and retrieval has not kept pace with the need of personnel in this field. This, therefore, has created opportunities for people who have the need for information storage and retrieval or document control thrust upon them. An analysis of the attendance at this conference would probably indicate that the number of chemists, biologists, clinicians, lawyers, administrators, and others greatly exceeds that of people trained purely in the information field. The history of the approach to literature control therefore most often reflects the thinking of technical personnel in fields other than the library sciences. Good fundamental bases are often wanting, too, because in new areas of endeavor the sense of need for perpetuity is often unrecognized. What begins as a fairly mundane and discrete problem, involving the effect of moisture on a unit of electronic equipment, may end up as an information center involving a number of lifetime

TABLE 1. *Chronology of information handling, Prevention of Deterioration Center*

Date(s)	*Sponsor–source*	*Information mobilization form*	*Content retrieval method*
Prior to 1944	Joint Army-Navy Deterioration Steering Committee of the National Defense Research Committee	No formal control of newly generated minutes of meetings, conferences and similar information. Outside reference collecting diffuse, uncentralized and personal to members of contributory committees	No known formal method; individual memory recall
1944–1945	Tropical Deterioration Information Center	1. Information Center miscellaneous reports	No known indexes or other formal retrieval method Some evidence of an incipient system of categories being considered
		2. Information Center reports	As above, two bibliographies in series of 14 reports, indexed, however, by subject
		3. *Tropical Deterioration Bulletin* No. 1–24A	Author and subject indexes issued 1946
		Tropical Deterioration Bulletin No.1–7A, Vol. II	Author and subject indexes issued 1946
		4. *Tropical Deterioration Bulletin* Supplements	Author and subject indexes issued 1946
		5. *Tropicalization News*	No known formal method
1945	Prevention of Deterioration Center, National Academy of Sciences (NAS)-National Research Council (NRC); Office of Research and Invention Contract	*Prevention of Deterioration Abstracts*	Abstracted and indexed approximately 1/3 of reports received annually, continued categorical divisions; balance not retrievable in any systematic way
1946	PDC, NAS-NRC; Office of Naval Research Contract	*PDC Abstracts*	Index to abstracted reports
1949	PDC, NAS-NRC; Office of Naval Research Contract	1. *PDC Abstracts*	Index
		2. Edge-punched card system installed for both abstracted and unabstracted reports. Described by Curtis Brown, 1955	Through use of a direct code on 14 separate categorical files
		3. Fungicide file established as an edge-punched card system	Indirect code based on modified American Chemical Society classification system
1956	PDC, NAS-NRC; Office of Naval Research Contract	Supplementary or satellite file of pertinent noncommercial specifications established	Indexed as an experimental group of documents prior to possible installation as a Peek-a-boo, field-punched card system
1957	PDC, NAS-NRC; Office of Naval Research Contract	*PDC Abstracts*, supplemented with extracts or telegraphic treatment of hitherto undisseminated reports	Conventional indexing supplemented with monoterm rubrics preparatory for subjecting information in Volume XV to Peek-a-boo treatment

careers correlating knowledge from numerous and not obviously related disciplines. Conversely, not every problem is blessed by a coincidence of events which fertilize it into a growing, self-sustaining, entity. If there is to be ultimately a universal system of information storage and retrieval, its roots must lie in the recognition that knowledge should never be subjected to the vagaries of indecision in terms of perpetuity of the organization doing the work. Each mobilization of knowledge, be it that represented by a student processing his first thesis or some international agency dealing with a universal problem in infinite time, should command the respect entitling it to its place in the universal system. Only this guarantee can be sobering enough to induce a proper approach to producing and processing each endeavor in a serious intussusceptive manner.

In Table 1 the outline of the development of information mobilization and recall methods at the Prevention of Deterioration Center is given. Examination of this table indicates that early in the history of this development the need for some control was recognized. The first elements of order will be found in the *Tropical Deterioration Bulletin* which divided the information into discrete areas of interest, reflecting in part the various committee organizations which were staffed separately by men from the textile, wood, leather, optical instrument, packaging, and similar fields. The biological aspect of the problem was conceded from the very beginning, and a section "Prevention of Growth of Organisms" under which subheadings of "Inorganic materials," "Organic materials-cellulosic" appeared, was established in 1944. A Table of Contents first appeared in April 1945, and this recognized the area of "Finished Assemblies," which included electrical and electronic equipment in addition to studies on optical instruments. This fundamental organization is still well reflected in the present 14 categories characteristic of the *Prevention of Deterioration Abstracts* today.

In 1955 Curtis A.Brown, formerly of the staff of the Prevention of Deterioration Center, described the retrieval methods used at the Center in his paper presented at the Minneapolis ACS meeting. Brown contributed to the study "Two Approaches to the Retrieval of Information from a Special Library" of which specifically, "Part I: Application of a Simple Punched-Card System to a Special Information Center," described Brown's interpretation of the then existing system. The possible application of a Uniterm Coordinate Indexing system to a segment of the same files described by Brown, was discussed by Mortimer Taube as Part II of the above study.

The Uniterm system approach to the Prevention of Deterioration Center task disclosed in a very practical way the most serious shortcoming of the punched card method as established at PDC. This is the fact that the edgepunched system is characterized by a limited number of descriptors, restricted to 1148 punching positions for the total system. On the other hand, an unrestricted or a maximum depth type of indexing as practiced by the uniterm approach which included specific materials, organisms, chemical names, and similar conceptual categories obviously was not handicapped by any fixed limitation of terms. The chosen documents generated approximately 3000 terms in the first 500 analyses with each document on an average contributing 25 uniterms. This volume yield of descriptors generated by only 2% of the total file, re-emphasized the capacity problem. Among the many attractive features of the uniterm approach, the most appealing was the abandonment of many of the restrictions, in terms of descriptors. The complete and nondiscriminate indexing as exemplified by the above trial in terms of descriptors or rubrics must be recognized, however, as a mixed blessing. The desire to get every possible concept from an article is inherent to a good indexer and is a commendable avarice. The indexing load as described by Taube was in no way embarrassing, in what we must consider a limited trial—limited being defined as the ratio of five hundred documents as compared with the total collection of 22,000 reports. Input time here, although not specifically defined in print, is probably greater than input time in terms of conforming to the fixed code of the edge-punched card. The advantages of deeper indexing need no champion. The limits of practicality, however, will not long be ignored, and posting through the supernumerary retrieval pathways bodes an increasingly greater burden. Assuming, too, that the 500 documents used in the trial were subjected to the posting procedure which involves dividing a card into ten vertical columns marked 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9 and the document numbers placed into these respective stalls by their terminal digit characteristic, the rigors of collating, again minimal with but the derivatives of 500 documents, undoubtedly becomes quite overbearing psychologically, if not physically, as the existing rubrics become associated with more and more documents. It is essential to remember that indexing per se is comparatively a light task and is only the basis for the far more onerous task of posting and collating. The mental image of a pretty young lady, or a sedate one for that matter, coping with the alignment of numerous cards, wherein she is to recognize collated documents by similar terminal digits in a veritable sea of numbers, is quite distressing. It is not so distressing, however, that the desirability of providing her with a more suitable read out device is not apparent. The end would seem to justify the means if the means were just a little less fatiguing.

It appeared then that the Center was on the horns of a dilemma—tempted by the fundamentally sound, deep indexing approach but intimidated by the overwhelming personnel problem and the pragmatic justification relative to embracing the "sin" of overengineering. Repeatedly, the edge-punched card, as used in a direct coding method, had been found to be too inadequate in its language capabilities. Originally conceived in a distinctive analytical way, the fundamental logic seemed comprehensive and sound. This involved seven cardinal philosophical considerations beginning with (1) *environmental agents* interacting with (2) *materials* of various types which are subject to (3) *deterioration* in different forms and degree thus in turn demanding various (4) *corrective* and *preventive considerations,* the value of which can be measured by (5) *test methods* which draw heavily on supportive (6) *applied* and *theoretical studies* on materials properties, mechanism of destruction, inhibition, and control, all of which come from established, contributory (7) *literature* consisting of reports, surveys, patents, bibliographies, etc.

The scope of the system was further enhanced very early in the plan by creating what is primarily a taxonomy of materials. This established the 14 satellite files in the constellation which was our PDC edge-punched card system. These satellite files are:

1. Biological agents and pesticides
2. Ceramics, cements, glass, and plaster
3. Electrical and electronic equipment
4. Lacquers, paints, and varnishes
5. Leather
6. Lubricants, fuels, and hydraulics fluids
7. Metals
8. Optical instruments and photographic equipment
9. Packaging and storage
10. Paper
11. Plastics, resins, rubbers, and waxes
12. Textiles and cordage
13. Wood
14. Weather and climate

Even with this studied attempt to gain versatility, supplemented further by the use of descriptive phrases and words written in on the card, there were too often times when the beast was both deaf and dumb. Gradually undedicated spaces were used to increase the vocabulary of the quasi-automaton, but the privilege of continuing to do this *ad infinitum* would, within the foreseeable future, be available no longer. The desirability of an open-ended system became very apparent. To add to this the very intriguing ability to collate, if and when the occasion demanded, was soon accepted by the Center as an evolutionary step to be recognized and abetted.

Independently of the Center, other investigators of the comparative merits of existing manual systems had been considering respective advantages and disadvantages, and when our edge-punched card system was selected as a guinea pig, we welcomed the exploration. Probably by way of comparative trial to evaluate extension or promotion of possibilities relative to their own needs, the Office of Naval Research offered a set of questions to both our system and the Uniterm system in 1955, and although the question-answering ability of both systems was sustained, no detailed analysis or comparison of the two methods has been published. Similar tests, however, in other areas with other collections have indicated that any comparable systems so tested usually give the proponents of the vying methods considerable thought-provoking data and indications of potential improvement areas, within the system each is championing. That this was true of our position became apparent after the comparative trial. What information we gained indicated that our performance in terms of question-answering ability was competitive. This was not satisfaction enough. Our growing sense of disaffection was not altered. The necessity to reword or interpret the questions, handling large numbers of cards, the feeling of insecurity concomitant with turning the system over to strangers less familiar with it than we, all helped to fix our determination to move into some more flexible manual information storage and retrieval method.

Improving the system by abandoning a direct code and substituting an indirect one was never seriously considered, primarily because we believed this would further isolate the information in terms of utility for untrained users, coders, indexers, and administrators. We suspected too, that dealing with an indirect code would have an unfavorable effect on handling time, thereby increasing costs.

The Prevention of Deterioration Center is most rigidly bound by the economics associated with a group cataloguing documents at about 200 a month. This average is rising, and the basic figures reported by Brown in 1955 no longer hold. At that time the increase per year was represented by approximately 2000 reports; however, in the last quarter of 1957 alone, 999 documents passed through the hands of our accession librarian. Table 2 gives a somewhat expanded statistical version of the accessions activity at the Center through the last four years. Peak years such as 1956, wherein several special studies were carried out are reflected in the number of documents involved. This indicates the inherent flexibility that any literature handling system must have, while recognizing the rigor of a fixed annual budget, staff, and facilities. We have accepted as an approximate measure of work load the tenet that processing a document for the storage and retrieval function is probably at least as time-consuming as conventional library cataloguing procedures. At no time have we considered abandoning the typical library practices of cataloguing and supplementary information control through title and author index approaches. Our desire relative to conventional cataloguing is to increase the specificity of the total information system by supplementing the catalog with an input and output system for more detailed information units. The function of the Center as a technical document lending institution is believed to be soundly founded in classical librarianship. Although the relationship between processing time and costs, comparing the library cataloguing of documents with information storage and retrieval handling, is not known, the cataloguing data are considered to be a realistic measure of work volume since at any given time, the ratio of handling times in each operation, cataloguing vs. information system input time, is a discrete number. The fluid condition in work methods relative to our evolving information storage and control methods makes this an unsuitable area for relative cost estimates, whereas the more standardized library procedures lend themselves better to accounting methods. For these reasons we generally measure our work load in terms of library rather than information system activity.

TABLE 2. Documents accession data, four-year period, Prevention of Deterioration Center

Year	*Ordered*	*Received*	*Catalogued*	
1954				
1st quarter	750	547	567	
2nd quarter	695	490	482	
3rd quarter	617	417	544	
4th quarter	728	435	379	
Total			1972	164/month
1955				
1st quarter	1048	798	512	
2nd quarter	1011	761	677	
3rd quarter	1302	881	604	
4th quarter	1110	871	505	
Total			2298	192/month
1956				
1st quarter	1022	769	836	
2nd quarter	1040	875	707	
3rd quarter	879	731	793	
4th quarter	846	691	541	
Total			2877	240/month
1957				
1st quarter	785	664	515	
2nd quarter	792	706	621	
3rd quarter	805	641	498	
4th quarter	999	790	612	
Total			2246	187/month
			Average	196/month

BASIC CONCEPTS IN AVAILABLE SYSTEMS

The stimulus of the comparison between two systems as indicated above served to crystallize many of the questions that had arisen from time to time as to what might constitute a better information storage and retrieval system. Here the dynamic flux of the state of this art was both an aid and a hindrance. In looking about for a tailor-made system that would service our needs, there were none found that could be accepted without reservation. The marginal punch-card system by now had been evaluated against time and, although fundamentally sound, was being found somewhat uncomfortably pinching in its limitations associated with capacity where no indirect code was involved. The need to handle a continually greater number of cards was proving a strain on the nervous and muscular system not only of ourselves, but even on some of the early proponents of this routine.[1] The limited capabilities of the edge-punched card in respect to collating information too had been placed on the balance and found lacking. An extension or revision of the edge-punched-card system seemed to be precluded. The change to a system involving high machine costs was considered only in passing and was only considered seriously from the aspect that ultimate growth of our system could conceivably demand, in the future, a machine method. Today the machine method ranks high in importance of consideration primarily because current practices in manual systems should be, in as much as it is conceivably possible, such that ultimate absorption into a machine system can be accomplished with optimal efficiency. The move from man to machine, at present is apparently governed by the physical and time limitations of man in terms of work volume. All other things being considered equal, in the several competitive systems available today, given an economical machine that would reduce the physical burden of handling an ever growing number of 5 by 8 edge-punched cards, this system would probably be ideal for certain type information system performance requirements. No allowances should be entertained, however, which will bar it from absorption into or interpretation by a universal approach.

Of the various possibilities that suggested themselves to us, all within the realm of manual methods, gradually the Batten or Peek-a-boo approach became our principal interest. Here there seemed to be numerous advantages which on the surface appeared to contribute to our aspirations for a more versatile, less trying method.

The major difficulty in the Peek-a-boo approach apparently resides, in common with other punch-card systems, in the demands associated with the art of indexing. Batten, pioneer in this field, in his appraisal of this method points out clearly the difficulties in this respect, stressing the fact that indexing was in his studies considered as an art and not a science, and with the passage of time others have pointed out that this continues to be the principal difficulty. J.W.Perry has succinctly summarized these difficulties in discussing aspect cards in the book *Documentation and Information Retrieval,* co-authored by him and Allen Kent.

The parallelism between needs at the Imperial Chemical Industry at the time

[1] J.W.Perry, personal communication.

of Batten's work in 1939 and patent literature control expresses well the difficulties common to many information groups and the Prevention of Deterioration Center. His analysis showed:

1. Inadequacy of classical indexing systems, particularly because they do not lend themselves to rejection of unwanted material by nontechnical personnel.
2. Fully mechanized systems are too expensive.
3. Classification pattern changes with time.
4. Hollerith (IBM) or Powers (Remington Rand) cards, suggest a fixed lace-work pattern for a given set of facts, but as the collection grows, the capacity of cards is exceeded and a costly reorganization and a retraining with the use of technical personnel become inevitable.

Reflective consideration of this problem finally led him to the conclusion that there was a need for a simple apparatus that would be open in the sense that it could absorb an infinite increase in items. The system should be capable of handling diverse material and permit fine classification from a great number of points of view. This suggested a concept of “one item, one card,” as opposed to one document for each card.

In the Batten system the information is considered and indexed from point of view of topic, class, subclass, subject, name, or property, and for each of these items a punchable card is provided. The card is divided into a large number of frames which are numbered by marginal coordinates. Each frame represents a patent “document.” As various aspects of each document are noted, the proper position on each card is punched. This process is continued until all the documents have been dealt with and the spaces on the card have been filled. This results in a series of cards perforated in such a way as to characterize the contents of each document. By superimposing selected cards one on the other in a stack, the visible perforation throughout the entire stack will be obtained. Wherever the frames represent patent specifications having the selected characteristics, the coordination or collation of the selected criteria or rubrics becomes apparent through the uninterrupted light path through coincident punches, thus yielding a tailored answer to a tailored question. In addition, this system provides answers of the yes or no, fit or not fit type question.

Batten notes further the special problems in indexing patents. He brings out the demanding exactness of the legal type of language, the necessity for having unambiguous terminology.

In the classification of the information in the patents, considerable thought was given to the sorting system which would be used throughout the procedure. It is strongly emphasized that no mechanical aids can compensate for a defective classification scheme, although the converse is true. There is a basic need for a stock set of characteristics to be used for coding. This set is subjected to revision as the need arises and consists primarily of additions to the catalog of rubrics. One of the characteristics of a good aspect is that it is readily combined with other aspects and is capable of expressing all the stages between a generic and a specific situation. It is to be noted that this approach is a contradiction to “fixed” types of index handlings which are mutually exclusive. Further, in the Batten system the classes and subclasses are selected on the principle that these types of aspects are to be anticipated and will be sought in combination. Classification of this kind is considered by Batten to be an art, and as such the creation of rules should be governed by the nature of the material and the objectives of the system. With plastics as an example, the chemical nature of the material, manufacturing processes, use areas, and invention details were chosen as indexing bases. The four selected criteria are considered to be reflective of the memory process by which the seeker should hope to find the material he needs. Expansion of these four points occurs only when necessary.

The apparatus used by Batten consisted of a movable spring-loaded punch in a jig. The original card size was 8 by 10 inch, divided into 400 frames. The limitation of this card, however, was that it covered only 400 items, and the subsequent card selected was the standard Hollerith card which permits 800 items and is adaptable to punching with a standard keyboard punch. It is to be remembered that this card was used in an inverse manner, i.e., each card represented a characteristic in a document, rather than the entire document.

At the time of Batten's report, published in 1951, the system had been in use over 7 years, and consisted of 6800 items. There are 14 series, 10 on the old and 4 on the new cards. These decks permit a chronological succession sorting of the catalogued material which would not be possible in a time random collection. The latest deck would represent the newest knowledge, the oldest deck the earlier knowledge.

Although Batten believed the system to be unknown generally, it appears that in France during World War II certain army service records and data had been manipulated by a similar system, and the work of American investigators seemed to be moving almost contemporaneously, in these same channels, independently of the British and French work.

This brief review of the early Peek-a-boo effort reveals a marked similarity in needs between those recognized by Batten and the needs of the Prevention of Deterioration Center.

In 1955 the general state of the art of manual systems was reviewed by I.A. Warheit. He points out that where specific demand requires detailed subject headings, the need for quick indexing and easy availability has been met by creating additional categories or subcategories, and the addition of sentence type modifiers to expanded indexes. Brown's paper on the system at PDC reveals that these expediencies were excercised in our case too.

The details of any expanded index, Warheit points out, require that the untrained library client have elaborate cross references in order to use the expanded system. Further disadvantages are that the entries are inconveniently on individual file cards, and any excessive wording makes them impractical to code or even to arrange alphabetically.

Currently, the sciences are being offered several basic systems involving the use of punch cards, both edge-punched and field-punched, and either hand- or machine-sorted. These are the edge-punched McBee and field-punched Peek-a-boo system and the method of coordinated indexing with uniterms. There are also some incipient systems which depend on magnetic or film storage technics involving possible use of cathode ray tubes, and which require expensive machines for their operation. These, however, are much too elaborate and expensive for use in an ordinary library or an information center, such as is illustrated by PDC.

Both the coordinated index and the Batten method depend on the utilization of a system of indexing which may be referred to as the post-combination or multiple-aspect approach, where individual items are indexed separately and later recombined to meet the needs of each specific search. The differences among the systems lie, in part, in the method of physically manipulating these subject heading components, or “monoterms.”

Reports, particularly those involving installed systems of uniterm coordinate indexing, were of interest to us for here at last seemed to be a method which would free us of punch holes. The uniterm coordinate indexing method takes into consideration the difficulties inherent to any punched-card system which we at PDC summarize as, primarily, mechanical problems—the system is too big, too heavy, too bulky, too inflexible, etc.—and any improvement lies in alleviating these conditions. The uniterm system approach to literature control takes the “too factor” into consideration and very neatly fractionates the punched card denoting many qualities by labeling ordinary cards with the terms indicating each separated quality. On the back of each card labeled with a concept, rubric, descriptor, aspect, heading, or what have you, and currently much in vogue is the phrase “unit term,” a list of numbers is assembled, representing references giving information on a rubric. The term rubric is the word long used by Dr. Jeanne MacCreight, formerly Index Editor at *Biological Abstracts,* and we wonder whether this might not have a claim to priority, at least over many of our newer words describing the indexing unit. Nevertheless, collation between two rubrics such as copper naphthenate and cordage in the uniterm system means comparing the backs of these cards for common report numbers posted thereon in 10 ranks. Adoption of this scheme frees one from gadgets—no needles, punches, or peekabooing is involved. Size limitations are presumably avoided, and no coding is necessary; it can handle an infinite number of documents without slowing down the searching process, and it is adaptable for machine operation in relation to posting and matching numbers. The system, however, still requires that precautions and controls of ordinary cataloguing be maintained, that thoughtful, serious indexing from parent documents, not abstracts, be practiced, that adequate uniterm control with an authority list be scrupulously maintained.

Warheit further analyzes costs, particularly in relation to conventional cataloguing, and concludes that coordinate indexing presents little or no economic advantages. Unfortunately, no comparison between a coordinate indexing method and the peekaboo system of information retrieval was noted. We hope to contribute through our own studies information which may be applicable in this evaluation.

The practical application of the uniterm coordinate indexing system has been reported by Jahoda as it pertains to the information control in a soap industry file. Here there are revealed several significant evolutionary developments and concessions requisite to reducing the method to practice. In essence the system bows to the need for a dictionary of terms based on the need to "minimize this dispersion of information in the index," which apparently occurs when synonyms, etc., are not reduced to unit concepts. The bound term need is recognized and compound terms are accepted into the system as their need becomes evident. Even with broad indexing as established for the system, auxiliary classified lists have been created. This indicates that, in the resolution of a practical approach to the problem of broad versus specific indexing, one system, admittedly desirable, cannot be efficiently and practically established by using the numbers matching technic for retrieval. Additional recall devices established include author card index, a chronological-serial number for primary sorting aided by different colored cards for various time periods, and classification of documents by prefixes indicating types such as trade literature, patents, photographs, and reprints. The burden of posting here seems self-evident and is listed along with false coordination and danger of physical loss of cards as being the three specific disadvantages of the system. The advantages lie, he found, in that the system provides intensive indexing, the file is compact, and its size is leveling off gradually, and it is a simple system fundamentally, if the basic input rules are enforced consistently. In discussing this paper with Jahoda, after its presentation, the impression arose that there was a growing apprehension relative to the burden of posting.

After considering these many facets contributing to our ability to make as proper a decision as possible relative to the collection at PDC, we were most strongly influenced by the Batten system developments as published on by Wildhack, Stern, and Smith of the Office of Basic Instrumentation (OBI), National Bureau of Standards. Development on this essentially Cordonnier-Batten[2] approach, renamed more functionally and descriptively Peek-a-boo by the team at OBI, consisted primarily in increasing the capacity of the field-punched card, instrumenting a precise, reproducible low-cost punch with appropriate gages and preparing an accurate read out device. The size of the punch was reduced to provide registration positions for 18,000 documents. The prototype machine has now been improved and several copies of the device are distributed among investigators for clinical trials. Through the courtesy of the OBI, PDC has one such machine for evaluation in terms of our file.

Prior to the use of this proposed new system, however, we were faced with the problem of readjusting our entire thinking on the basis of now having available to us a potentially adequate and open end system containing no restrictions in terms of a limited code. The penultimate, we considered, was a hypothetically ideal number of index entries per document. The physical limitations, however, could not be readily assessed. No staff increases were contemplated. Although coding as required by the older system would no longer be necessary, no information was available on how our coding and indexing load would be related to our document accession load. Out of deference to our previous indexing method used in the *Prevention of Deterioration Abstracts* our conventional approach we felt had to be continued. This contains the usual element of precoordination as interpreted by our indexing-abstracting staff, based on each individual's own understanding of the document. Although we considered a simple listing of the rubrics generated by each document from which it was conjectured that our readers could do their own collating as required, the mechanical problems in creating such a coordinate index apart from a card system did not appear to be readily overcome. Theoretical considerations of adequate indexing were difficult too in all terms of the absolute. How should the index discriminating sense be judged? Should every trade name, every possible organic compound, all genera and species of organisms become rubrics? After much discussion it was conceded that possibly a principle of self-limitation would become operative in our field. We considered that not all possible organic compounds will ever fall into the sphere of a single system, nor will all the organisms in the world be associated, for example, with the field of biological deterioration. It was recognized, therefore, that one of our early functions would be to try to measure in some way the size of the vocabulary

[2] Inventorship is not implied here.

inherent to our business and to find at which point, if at all, the principle of self-limitation becomes operative. From looking at other systems we were inclined to believe that this is a manageable number of rubrics and that the total number is not one by which we should be intimidated. Taube's work on our files showed that 500 reports generated 3000 uniterms, but this gave us no indication at which point the leveling off would occur. Taube's original estimate of what our total vocabulary would probably be was revised upward subsequently, too, based on the prolific yield associated with the first 500 documents.

EXPERIENCES IN INDEXING A SATELLITE FILE

At this point one of our former staff editors, William Wood, suggested that our collection of hitherto unindexed United States specifications would be a good experimental file. This suggestion was recognized as an opportunity to crystallize some of the physical and ideological difficulties more realistically than a theoretical evaluation, and it was adopted.

Indexing the specifications file produced several useful items of information. First, once an individual has indexed conventionally, that is, by using descriptive phrases and precollating with the document in hand, there is a marked psychological block to reducing the content of a document to a series of disjointed words. The monoterm selection type of indexing does not, apparently, produce the same satisfaction that comes with the correlating of ideas presented by an author and as interpreted by an indexer. This judgment sense, ability, call it what you may, is strongly reminiscent of what by many people is recognized as the art of indexing. An indexer prides himself on the fact that he is capturing ideas and concepts, not merely herding words into a corral, which later can be reassembled at will to produce compound concepts so meticulously before disassociated, or perhaps even to give interpretations which may be entirely unrelated to parent material. The meaning of words alone is basis for controversy, collating out of context may well be nothing more than superimposing two or more variables, the selected rubrics in this instance, so that a result reminiscent of events at the Tower of Babel might ensue. Even currently in our work, apart from time limitations, there is a marked reluctance to take out of context the various single or bound terms that are characteristic of a document. Theoretically it is possible to argue, sincerely and convincingly, that any information recall system which recognizes the value of collation can function ideally only when each document yields its full number of bits. We are forced, by practicality, however, to recognize this again as a depth of indexing problem and thereby grant our indexer the license to use his or her judgment in relation to any given bit. Even though a casual reference has often sparked a significant series of investigations or provides the missing link in a much wanted answer, we concede that the laws of probability militate against such an ideal coincidence. The use of a dictionary of terms is often proposed to minimize the necessity for decision making; however, it must be remembered that although the dictionary of terms will definitely enumerate the words in your technical vocabulary, it cannot help in judging significance. For example, no dictionary of terms can direct on how he should deal with this example. A report comes to PDC, and it deals with a specially prepared, waterproofed, treated paper which in experimental trials yields a great deal of good data indicating that the subject paper is quite worthy of application trials in certain packaging uses. In conclusion, the author summarizes the qualities of the wrapping paper and emphasizes that certain lumber shipments can very well be protected from the effects of moisture in possibly the hold of a ship. Hopefully, too, the proponent suggests, in passing, that of course this wrapping paper should also be protective in relation to dust. There are no dust protection studies reported. Dust is certainly by its abrasive qualities, its soiling properties, and its microbiological implications truly a rubric in the vocabulary of a deterioration prevention information center. Time demands and quality of factual information standards as interpreted by the analysts, even though there are arguments both pro and con for inclusion or exclusion of this experimentally untested reference, govern the final disposition. The ultimate decision, we have found, rests in the hands of our abstracting-indexing group, who are guided primarily by their own interpretation of each document they survey.

The indexing of our satellite group of specifications produced from 2 to 24 entries for each specification. The indexing of the specifications provided us with the opportunity of assessing the applicability of each document to our field of endeavor. Although 2500 pieces were examined, far from all were pertinent, and the finished index produced only two index entries for each specification.

The specification index, we believe, produced a conservative number of characteristic terms because specifications are, in their physical conceptual approach, infinitely more standardized than the random documents that come to our files. The concentrated indexing effort produced from zero to 24 entries for each document, and the final average number proved to be two. In considering this possibly low value, we feel we would need to remember that the specifications are selected by title only and often prove minimal in contributory value. Specifications, too, are very uniform and standard in format and thus minimize a diversity of entries. Our indexer, who had long worked with specifications, understood that each specification usually includes characteristic categorical concepts such as scope, applicable related documents and drawings, requirements, sampling, inspection and test procedure, preparation for delivery, packaging, and notes. The indexing therefore was more strongly oriented toward materials and methods, typically the prevention of deterioration point of view, rather than toward a more comprehensive type of registration. Approach to the specifications file implies understanding the categorical makeup, and queries put to the collection would be centered on the need for more specific information. The orientation of indexing toward materials and methods proved fortuitous in two ways.

Primarily the index to the specifications placed a great deal of hitherto unavailable information into the overall system of information resurrection, and in a short while after its completion, it proved useful in several instances, enabling us to make a comprehensive search in several areas where, within the realm of specification requirements, hitherto we had been progressing in an unsatisfactory manner. Secondarily and more importantly, however, and we believe this to be a fundamental principle which will continue to be of service to us, we discovered for ourselves what we have chosen to call the principle of re-entry. This may be defined as the concept which gives an indexing system the potentiality of leading the searcher back to cardinal entries for the purpose of developing a chosen area in a more definitive way. Throughout the history of information storage and retrieval at the Prevention of Deterioration Center, we have studiously avoided, and, in fact in the edge-punched card system, made little allowance for specific names of organisms, compounds, trade names, and similar large and comprehensive areas of information, primarily because the capacity of the systems available to us then did not encourage an unlimited vocabulary. Secondly, the economic justification for a system encompassing a large and possibly uneconomic factor of redundancy and the candid recognition that it is very unsatisfactory to try to anticipate every question which may or may not be put to the system forced us to accept the limitation, barring direct entry of specific organisms, experimental compound names, and similar categories which we anticipate in part are best handled in separate satellite collections. Our respect for the possible need of a hierarchical approach encouraged us, too, to try for balanced indexing in a taxonomic sense, since it appeared to us to be a poor indexing approach to analyze subject material in certain taxonomically developed areas to a finite degree whereas in other undeveloped taxonomic regions no similar degree of indexing seemed readily apparent. Whether we have overemphasized the value and significance of a hierarchy in information handling is debatable, but with limited funds and the need to devise a useful, practical economic system, we have been unable to dismiss considering the hierarchical approach as an instrument of balance aiding us in striving for a uniform depth of indexing. A uniform depth of indexing by experienced people plus the use of the re-entry principle gives us a sound and economic system in which the cost of redundancy is kept at a minimal value. No re-entry into the files is made until the need arises, and the cost of the re-entry is then amortized against the new need.

The soundness of the re-entry principle was made manifest to us in using the specifications file shortly after its completion. In considering the applied microbiology that is written or implied in the specifications concerning themselves with the deterioration of materials, it became desirable to catalog the organisms associated with all the tests covered in these documents. By pulling all entries under organic materials within the established fields of interest of the center, e.g., objects made of leather, paper, wood fabrics of various kinds, and coatings, a comprehensive roster of organisms associated with testing the aforementioned materials was made available. This, when reduced to index entries, is easily assimilated into the file for future needs, and it satisfied the need which engendered and financed the request. Development of the re-entry principle will require comprehension and obedience to existing good indexing practices plus possible subjective operational concepts based on the aims and purposes of each individual collection. Since a universal system of literature control continues to be quite elusive and applied information centers are required to store and release on call vast amounts of knowledge, always at minimal cost, we feel that the principle of re-entry with its clear-cut cost accountability is a concept of considerable interest and value to us. It seems a logical approach to the problem of redundancy.

PRACTICAL CONSIDERATIONS IN ESTABLISHING THE FIELD PUNCHED-CARD SYSTEM

During the experimental period prior to the adoption of the field punched coordinate system, which centered on our work with the hitherto unindexed specifications file, the problem of establishing a dictionary of terms required resolution. Our attitude on the desirability of having beforehand a comprehensive dictionary of terms has been that, in essence, this would be a real help in our current endeavor. As in many other systems, the field punched-card coordinate method requires the preparation of a standard dictionary of terms. It has been estimated that for our field this dictionary would consist of about

5000 to 8000 terms. Once established, this dictionary must be adhered to strictly. For each term in the dictionary, one card is dedicated. We have in mind cards 5 by 8 inches, providing sufficient space to dedicate 18,000 small holes. Each hole represents the number of a report. Thus, when 18,000 reports have been received and coded, it is necessary to start a new deck of cards for the next 18,000.

In brief explanation of the actual working system, suppose we receive a report on the subject of copper naphthenate being used as a fungicide to protect a cotton textile against mildew in the tropics. Let us say that this report is number 12,500. Cards corresponding to the terms copper, naphthenate, cotton, textile, fungicide, mildew, and tropics are removed from the file, and hole number 12,500 is punched. Let us also suppose that this same operation is carried out for a large number of reports with the standard dictionary of terms. Later, when a request is received for information on the use of copper naphthenate as a protectant of cotton fabrics against mildew in the tropics, the cards corresponding to the dictionary terms in the request are pulled from the file. The cards are aligned in a scanning or read out device containing a light source. Holes common to all cards will thus permit light to shine through and indicate the numbers of reports containing information common to the dictionary terms of the request. Consideration of this example illustrates that either of two courses relative to the time at which a dictionary of terms can or must be made available to the system may be taken. We considered seriously establishing the dictionary before our indexing. Our enthusiasm, however, in wanting to wade into the new technic plus some rationalizing on the basis that the current literature indexing would produce more synonyms, homonyms, and suggestions for arbitrary definitions than we could possibly recall led us to accepting the working practice of building the dictionary of terms and concept ranking as we moved ahead with our work. This is open to challenge from the efficiency point of view, particularly since with our accumulation of abstract indexes, index to the textbook *Deterioration of Materials,* edited by Greathouse and Wessel, and general group familiarity resulting from over ten years of Center operation we feel that the dictionary of terms does exist, in essence, within the four walls of the Center. Collecting this information into a discrete index is, however, considered to be but a matter of time, because after processing just five issues of *Prevention of Deterioration Abstracts* within the newer framework of accepting terms with a minimum of restrictions, the dictionary is growing well in structure and depth. The need for absorbing concepts from our older sources may be unnecessary. Editing of the newly acquired terms with proper signification of deletions, cross-indexing, see also entries, etc., we believe will keep our system useful, comprehensible, and contemporary. This self-generated collection of descriptors will need, we believe, little more editing than any collection of raw terms.

The growth of our vocabulary in the category of metals and plastics at the conclusion of the fifth issue of Volume XV is presented in Tables 3–8. We do

TABLE 3. Quantitative generation of keywords in "Metals" category. Issues 1 to 5 inclusive, Volume XV, Prevention of Deterioration Abstracts

Number of reports (approximately 100 sheep[a] and 90 goats)	190
Number of keywords generated	432
Total number of entries	1373
Average number of entries per report	7.2

[a] "Sheep" are abstracted reports; "goats" are non-abstracted reports.

not present these data as a significant sample from which to make any valid extrapolations. We do hope eventually, with more of this type of data, to ascertain our plateaus and uncover corollary information which may be of use in general application areas for information groups collecting and analyzing

TABLE 4. Frequency of keyword occurrence, category "Metals," Issues 1 to 5 inclusive. Volume XV, Prevention of Deterioration Abstracts

Frequency of occurrence	*Number of key words*	*Frequency of occurrence*	*Number of keywords*
1	257	17	2
2	76	19	2
3	21	20	2
4	7	21	2
5	13	22	2
6	24	24	1
7	7	25	1
8	1	29	1
9	3	36	1
10	1	37	1
11	2	60	1
13	3	78	1
15	2	96	1
16	1		

data in an information center accessing documents at a rate comparable to the rate at PDC. Though one frequently finds general statements referring to the estimated approximate size of a vocabulary in any given area, precise, developmental,

TABLE 5. *Keywords and frequency of occurrence in the category of "Metals," Issues 1 to 5 inclusive, Volume XV, Prevention of Deterioration Abstracts*

Keyword	*Entries*
Acetylation	1
Acids	1
Acrylonitrile rubber	1
Adhesion	1
Adsorption	1
Aeration	1
Aging	1
Agricultural	1
Aircraft	7
Alkalies	1
Alkalinity	1
Allotropy	2
Alloy	1
Alloying	1
Alloying additives	2
Alloying agents	7
Alloying metal (element, additive)	2
Altitude	1
Aluminum	36
Aluminum alloy(s)	17
Amines	3
Anodes	6
Anodes (sacrificial)	1
Anodizing	5
Anticorrosive additives	16
Anticorrosive treatment(s)	5
Antifouling	1
Antifreeze	1
Antitarnish	1
Application(s)	9
Aqueous solution(s)	2
Asbestos	1
Asbestos-cement	1
Atmosphere	1
Atmospheric pollution	1
Austenitic	1
Automotive engines	1
Awnings	1
Bacteria	2
Bactericides	1
Ballast	1
Batteries	1
Benzophenones	1
Beryllium	1
Bibliography	1
Bis(5-chloro-2-hydroxyphenyl)-methane	1
Bitumens	3
Boilers	1
Bolts	1
Boron	1
Brass	4
Bridges	1
Brine *See* Salt water	
Bronze	1
Butadiene rubber	1
Butyl rubber	1
Cable(s)	5
Cable sheath	1
Cadmium	2
Calcium silicate	1
Cans *See* Containers	
Carbon black	2
Carbon dioxide	1
Casings *See* Shields	
Casting	1
Cathodic protection	24
Cavitation	1
Cellulose resins	1
Cement	2
Ceramics	1
Cermets	1
Chemical	1
Chemical composition	1
Chemical plating	1
Chemical resistance	2
Chemical structure	2
Chemicals	2
Chlorinated rubber	1
Chlorination	1
Chlorine	1
Chlorophyll	1
Chromallizing *See* Chromizing	
Chromate(s)	9
Chromate treatments	1
Chromic acid	2
Chromium	8
Chromium alloys	1
Chromium compounds	1
Chromizing	3
Cladding	2
Climate	4
Coal tar	2
Coatings	29
Cobalt alloys	2
Coloration	1
Columbium	1
Combustion products	1
Commercial	2
Commercial paper	4
Commercial publication	1

Keyword	*Entries*
Communication systems	1
Compendium	3
Compilation	1
Components	2
Compound 1080	1
Concrete	4
Condensation	1
Condenser tubes	1
Container(s)	2
Contaminants	5
Conversion coating *See* Anticorrosive treatments	
Cooling systems	1
Cooling towers	1
Copper	19
Copper alloys	7
Copper compounds (organic)	1
Copper-8-quinolinolate	1
Corronizing	1
Corrosion	78
Corrosion inhibitors	2
Corrosion resistance	7
Corrosion resistant	2
Cost(s)	10
Cotton	1
Couples	1
Coupling	3
Cracking	1
Creep	1
Creosote	1
Crevice corrosion	1
Crystallography	2
Cycling	1
Cylinders	2
Decking	1
Deformation	1
Dehumidification	2
Design	5
Detergents	1
Dichlorophene	1
Dielectric failure	1
Diesel engines	1
Discontinuities	2
Douglas-fir	1
Dow *See* Anodizing	
Duck	1
Dust	2
Dyes	1
Economics	1
Electrical contacts	1
Electrical drainage	1
Electrical resistance	1
Electrode potential	2

Keyword	*Entries*
Electrolytes	3
Electroplate	5
Electroplating	5
Enamels	1
Encasements	1
End uses	3
Engines	3
Epoxy resins	2
Equipment	1
Erosion	2
Exhaust gases	1
Fabrication	1
Farm equipment	1
Fastenings	2
Fatigue	2
Fenders *See* Shields	
Ferricyanides	1
Ferrous metals	21
Fertilizer	1
Film(s)	5
Flame spraying	2
Fluorine compounds (inorganic)	1
Foreign	13
Fretting corrosion	1
Fuel ash	1
Fuels	1
Fungi	1
Fungicides	1
Galling	1
Galvanic corrosion	2
Galvanized coatings	1
Galvanizing	2
Galvanizing *See* Zinc	
Gas lines	1
Gases (liquid)	1
Gasoline	1
Germany	1
Glass	2
Glass textiles	1
Gophers	1
Graphite	1
Grease	1
GR-S	1
HAE *See* Anodizing	
Halogen compounds	1
Halogens	1
Hardness	1
Hardware	1
Heat transfer	1
High pressure	1
High temperature(s)	20
Holidays *See* Discontinuities	
Hot water	1

Keyword	*Entries*
Houses	1
Humidity	2
Humidity (water vapor)	4
Hydrochloric acid	2
Hydrogen peroxide	1
Hydrogen sulfide	2
Impurities	1
Inconel	1
Industrial	3
Inspection	1
Insulation	1
Insulation (heat)	1
Intergranular corrosion	2
Intermetallics	1
Iridite *See* Anodizing	
Iron	11
Iron alloy(s)	2
Iron oxide(s)	2
Irrigation	2
Israel	2
Joining	1
Joints	1
Kanegen *See* Nickel alloys	
Kerosine	1
Lacquers	1
Laminates	3
Lead	5
Lead alloys	3
Light	1
Light stabilizers	1
Lime	1
Limnoria	1
Linings	2
Lithium	1
Low temperature(s)	5
Magnesia	1
Magnesium	13
Magnesium alloy(s)	7
Magnetic	1
Maintenance	4
Marine	7
Marine borers	1
Marine fouling	1
Mechanical properties	2
Mechanism(s)	15
Mercury compounds	1
Metal(s)	19
Metallizing	2
Microstructure	1
Migration (silver)	1
Mollerizing	1
Molybdate	2
Molybdenum	3

Keyword	*Entries*
Monel	1
Naphthenate	1
Neoprene	2
Nephroe *See* Nickel alloys	
Nickel	8
Nickel alloys	5
Nomenclature	2
Nuts	1
Oil(s)	4
Oil resistance	1
Okun *See* Solder or putty	
Oxalate	1
Oxidation	7
Oxide(s)	6
Oxidizing agent	1
Packaging	6
Paint(s)	6
Paper	2
Passivation	6
Patent	11
Permeability	1
Pertechnetate	2
pH	2
Phenolic resins	2
Phosphate	2
Phosphates	3
Phosphorus compounds (organo)	1
Photoactivity	1
Piers	3
Pigment(s)	6
Piling(s)	5
Pipe(s)	20
Piston guides	1
Pistons	1
Pitting	3
Plastics	2
Platinum	1
Poisons	1
Poland	1
Polarization	1
Poles	1
Pollution	1
Polyester resin	3
Polyethylene	2
Polyvinyl chloride	1
Potting resin	1
Powder	1
Pressure treatment	1
Primer	1
Progress report	2
Putty	1
Pyridine derivs.	1
8-Quinolinol	1

Keyword	*Entries*
Radiation	1
Railway equipment	1
Rain	1
Reactors	3
Reducing agent	1
Reinforcement(s)	2
Relaxation	1
Research	96
Resin	1
Resistivity	1
Review	60
Rivets	1
Roofing	1
Rubber	2
Rubber (synthetic)	1
Salt solutions	1
Salt spray	2
Salt water	1
Scale	1
Scaling	1
Sea water	15
Sealing	1
Seizing	2
Service life	1
Sheathing	1
Sheaths	1
Sheet	1
Shields	1
Ships	5
Shrouding	1
Silicon	3
Silicones	1
Siliconizing	2
Silver	6
Silver alloy	1
Slushing compounds	2
Sodium benzoate	1
Sodium hypochlorite	1
Sodium nitrite	2
Soil	10
Soil corrosion	1
Solder(s)	3
Solutions	1
Solvents	2
Sound Waves	1
Specifications	1
Sprays	1
Stabilizers	1
Stain	1
Stainless steel	25
Stannic chloride	1
Statistics	1
Steam	1
Steam *See* Water and High temperature	
Stearic acid	1
Steel	37
Storage	2
Stray currents	1
Stress	2
Stress corrosion	7
Stress cracking	1
Structural materials	1
Structures	2
Strychnine	1
Sulfur	1
Sulfuric acid	2
Surface preparation	2
Surface treatments	7
Tabulation	1
Tank cars	1
Tankers	3
Tantalum	1
Tape(s)	3
Tarnish	1
Tarnish *See* Stain	
Temperature	1
Termites	1
Ternary systems	1
Test(s)	21
Test equipment	6
Textiles	1
Theory	13
Thermosensitive	1
Thiokol rubber	1
Thorium	1
Thread	1
Tin	5
Tin-lead	1
Titanium	9
Titanium alloys	1
Tow reels	1
Transport	1
Transportation	1
Tropical	2
Tropics	1
Tubing	2
Tungstate	2
Ultraviolet light	1
Underground	6
Urban	2
Urea-formaldehyde resin	1
USSR	1
Utensils	1
Vanadium pentoxide	1
Vapor phase inhibitors	1

Keyword	*Entries*	*Keyword*	*Entries*
Vat dying	1	Weather resistance	1
Vehicles	1	Weathering	2
Velocity	1	Welding	2
Vessels	1	Welds	2
Vibration	2	Whiskers	1
Villanova Cycling Machine		White rust	1
See Test equipment; corrosion		Wire	2
inhibitors		Wood	2
Vinyl chloride	[illegible]	Wood preservatives	1
Vinyl chloride polymers	1	Wrappings	1
Vinyl resins	3	Wrought metal	1
Vulcanizing agents	1	Zinc	22
Water	17	Zinc chromate	1
Water systems	1	Zinc compounds (organic)	1
Waterproofing agents	1	Zinc oxide	1
Wear	2	Zirconium	1
Weather	1	Zirconium alloys	2

TABLE 6. Quantitative generation of keywords in "Plastics" category. Issues 1 to 5 inclusive, Prevention of Deterioration Abstracts

Number of reports (Approximately 67 sheep and 49 goats)	114
Number of keywords generated	316
Number of keywords generated per 100 reports	277
Total number of entries	815
Average number of keywords per report	7.15

TABLE 7. Frequency of keyword occurrence, category "Plastics." Issues 1 to 5 inclusive, Volume XV, Prevention of Deterioration Abstracts

Frequency of occurrence	*Number of keywords*	*Frequency of occurrence*	*Number of keywords*
1	196	11	3
2	45	13	1
3	12	14	3
4	13	17	1
5	8	18	1
6	4	19	1
7	7	27	1
8	4	30	1
9	2	49	1
10	3		

TABLE 8. *Unedited keywords and frequency of occurrence in the category of "Plastics." Issues 1 to 5 inclusive, Volume XV, Prevention of Deterioration Abstracts.*

Keyword	*Entries*
Abrasion	1
Abrasion resistance	1
Absorption	1
Acids (organic)	1
Acrylic rubber	1
Acrylonitrile polymers	3
Adhesion	1
Adhesive(s)	7
Age resistance	1
Aggregates	1
Aging	9
Aircraft equipment	1
Aluminum silicate	1
Amines	1
Ammunition	1
Animal glue	1
Anticorrosive additives	1
Anticorrosive coatings	3
Antimony compounds (organic)	1
Antioxidant(s)	10
Antiozonants	10
Antiozonants (candidates)	1
Application	2
Arctic	1
Asphalt	1½
Atomic radiation	1
Automotive components	2
Autopolymerizing	1
Bacteria	1
Bactericide	1
Belts	1
Benzoic acid	1
Bibliography	1
Birch	1
Blooming	1
Book	2
Brittleness	1
Butyl rubber	4
Butyl rubber, brominated	1
Cables	1
Calcium carbonate	1
Camphor	1
Carbon black	2
Casein	1
Cellulose	1
Cellulose acetate	1
Cellulose acetate butyrate	1
Cellulose resins	1
Chemical modification	1
Chemical reactions	1
Chemical resistance	10
Chemical structure	2
Chlorosulfonated polyethylene *See* Hypalon	
Chlorotrifluoroethylene *See* Fluorothene	
Cladding	1
Coatings	11
Color	1
Commercial	5
Containers	1
Corona	1
Corrosion resistance	1
Cost	1
Cracking	3
Creep	1
Curing	2
Degradation	1
Depolymerization	1
Desert	1
Diester oils	1
Dimensional stabilization	1
Discoloration	1
Douglas fir	1
Durability	1
Dynamic properties	1
Economics	1
Elastomers	14
Electric insulation	1
Electrical properties	2
End use(s)	4
Epoxy-polysulfide resin	1
Epoxy resin(s)	7
Esters	1
Extenders	1
Extreme temperature	1
Extrusions	1
Fabrics	1
Fasteners	1
Ferrous metals	1
Fillers	5
Film(s)	5
Fire resistance	1
Fireproofing	2
Fittings	2
Flexibility	1
Flexural strength	1
Fluorine compounds (organo)	6
Fluorocarbons	2
Fluoroelastomers	1

Keyword	*Entries*	*Keyword*	*Entries*
Fluorosilicone rubber	1	Microorganisms	1
Fluorothene	1	Migration	1
Foam(s)	2	Moldings	1
Formic acid	1	Molds	1
Formulation	2	Neoprene	3
Fortisan	1	Nitric acid	1
Fuel resistance	4	Nitrile rubber	4
Fungi	3	Nitrogen dioxide	1
Fungitoxics	1	*p*-Nitrophenol	1
Fungus resistance	1	Nomenclature	1
Fungusproofing	1	Nutrition	1
Gaskets	1	Nylon	5
Gasoline	1	Oak	1
Glass	1	Oil resistance	8
Glass fabric	2	Olefins	2
Glass fiber(s)	2	Oxalic acid	2
Glossary	1	Oxidants	1
Greaseproofing	1	Oxidation	5
GR-S	13	Oxygen	2
High temperature(s)	27	Ozone	11
High temperature resistance	1	Ozone cracking	1
Humidity	4	Ozone resistance	4
Hycar	1	Ozonization	1
Hydraulic fluids	1	Packaging	2
Hydrolysis	1	Paintings	1
Hypalon	5	Paper	3
Inert atmospheres	1	Patent	18
Infrared	2	Pentachlorophenol	1
Inhibitors	1	Permeability	1
Interface agents	1	Phenol-formaldehyde resin	1
Irrigation ditches	1	Phenolic resins	6
Isoprene	1	Phenols	3
Joints	2	Phosphorus compounds (organo)	1
Jungle	1	Photochemistry	1
Laminates	7	Photodegradation	1
Leather	2	Photometry	1
Leather substitute	1	Pigments	1
Light	4	Pine	1
Light resistance	1	Pipe	7
Light stabilizers	1	Plasticizers	4
Liners	1	Plastics	9
Low temperature(s)	6	Plywood	2
Lubricants	1	Polyester(s)	1
Magnesia	1	Polyester resins	3
Marine	1	Polyethylene	14
Mathematics	1	Polymerization	1
Mechanical properties	4	Polymers	8
Mechanisms	4	Polystyrene	1
Melamine-formaldehyde resin	1	Polysulfide copolymers	1
Melamine resin	1	Polysulfide polymers	2
Metal halides	1	Polysulfide rubber	1
Metals	2	Polysulfide-silicone polymers	1
Microbicides	1	Polytetrafluoroethylene	2

Keyword	*Entries*
Polyvinyl chloride	7
Polyvinylidene chloride	1
Potting compound	2
Primers	1
Process	1
Progress report	17
Pumps	1
Radiation	2
Radiation (nuclear)	1
Reinforcement	2
Research	49
Residues	1
Resins	2
Resorcinol-formaldehyde resin	1
Resorcinol resins	1
Review	19
Rheology	1
Rodents	1
Rubber	30
Rubber (chlorinated)	2
Rubber hydrochloride	1
Rubber (raw)	1
Salicylic acid	1
Salt water	1
Screening	1
Sea water	1
Seals	2
Shellac	1
Shock resistance	1
Silastic *See* Silicone rubber	
Silicone copolymers	1
Silicone-polysulfide polymers	1
Silicone rubber	3
Silicones	5
Siloxanes	1
Smoke	1
Sodium benzoate	2
Specifications	2
Spectroscopy	1
Splicing	1
Stability	1
Stabilizers	7
Staining	1
Standards	1
Steel	2
Storage	3
Stress	1
Stress relaxation	1
Structures	1
Styrene	1
Styrene-butadiene copolymers	2
Styrene-butadiene rubber	1
Styrene resin	1

Keyword	*Entries*
Sulfides	1
Sulfur	2
Sulfur compounds (organo)	1
Sulfuration	1
Summary	1
Swelling	1
Symposium	2
Synthesis	11
Tanks	1
Tar	1
Techniques	1
Teflon *See* Polytetrafluoroethylene	
Temperate	2
Temperature	3
Temperature extremes	5
Temperature resistance	1
Test(s)	14
Theory	1
Thiadiazole (derivs.)	1
Thiokol	2
Thymol	1
Tin compounds (organo)	4
Tire(s)	8
Tire treads	1
Triallyl cyanurate	1
Tropical	1
Ultraviolet	3
Underground	1
Urea (derivs.)	1
Urea-formaldehyde resin	1
Urea resins	2
Urethan	1
Urethan polymers	1
Urethan rubber	1
Vacuum	1
Vinyl chloride	2
Vinyl chloride polymers	2
Vinyl chloride resin	1
Vinyl halide resins	1
Vinyl resin(s)	4
Viscosity	1
Vulcanizing agent(s)	6
Water	2
Water vapor	2
Waterproofing	4
Wax	7
Weapons	1
Wear	1
Weather	8
Weather resistance	1
Windshields	1
Wood	2
Wood preservatives	1

quantitative information is not readily available. Our establishment of the field punched-card coordinate system with its mustering in of rubrics gives us the opportunity to do this for our Center. General applicability may follow.

It is pleasant to realize that already our descriptive vocabulary is far superior to our previous code. The truer reflection of our analysis is especially satisfying because to date we have not had to increase staff and adoption of the system has enabled us to include telegraphic extracts of all retained documents and to provide our information seekers with a service far more comprehensive than any hitherto offered.

ACKNOWLEDGMENTS

Sincere thanks are expressed to Mrs. Grace Chapman, Supervisor, *Prevention of Deterioration Abstracts,* for data appearing in Tables 3 to 8 inclusive, to Miss Faith Bissell, PDC librarian, to Mr. William J.Wood, Jr., formerly Editor at the Center, for his indexing accomplishments in the specifications file, and to the entire staff for their valuable assistance in preparation of this report.

BIBLIOGRAPHY

BLOOMFIELD, MASSE. Evaluation of coordinate indexing at the Naval Ordnance Test Station. *Am. Document. 8,* 22–25. (1957).

BROWN, CURTIS L. Two approaches to the retrieval of information from a special library. Part I. Application of a simple punched-card system to a special information center. PEAKES, GILBERT L., ALLEN KENT, and JAMES W.PERRY, Editors, *Advances in Documentation and Library Science,* Vol. 1: *Progress Report in Chemical Literature Retrieval.* Interscience, New York, 1957, pp. 121–142.

BROWN, WILLIAM FULLER, JR., and GLEN ONEAL, Jr. Library searches with punched-card machines. *Science 123,* 722–723 (1956).

CASEY, ROBERT S., and JAMES W.PERRY, Editors. *Punched Cards: Their Applications to Science and Industry.* Reinhold, New York, 1951.

CONNOR, JOHN M. The need for documentation to government specifications. *Special Libraries 47,* 152–155 (April 1956).

Documentation, Inc., Washington, D.C. Application of the uniterm System and the Association of Ideas to a Special Library File, by Mortimer Taube. August 1955. (Technical Report 10).

FINDLAY, W.P.K. The use of perforated cards for preliminary identification of fungi. *Brit. Mycol Soc. Trans. 31(1/2),* 106–111 (1957).

GAMBLE, DEAN F. A coordinate index of organic compounds. (Paper presented before Chemical Literature Division, 127th Meeting American Chemical Society, Cincinnati, Ohio, March 31, 1955.)

GARRETT, G.T., and O.OSMON, Jr. New punched card system will help you organize corrosion data. *Chem. Eng. 64(6)*, 342, 344, 346, 348 (1957).

GULL, C.D. Implications for the storage and retrieval of knowledge. *Library Quart. 25,333–343* (1955).

MATHAY, W.L., and R.B.HOXENG. A classification and filing system for corrosion literature. *Corrosion 12*(11)*,* 72–76 (1956).

Microcite, an aid to more effective referencing. *Nat. Bureau of Standards (U. S.) Tech. New Bull. 41,* 141–142 (1957).

Multidimensional indexing speeds instrument-document search at the National Bureau of Standards. *Chem. Processing 19*(4)*,*96–97 (1956).

NELSON, W.L. Technical filing system. *Oil Gas J. 54*(25), 111 (1955).

OPLER, ASCHER, and TED R.NORTON. New speed to structural searches. *Chem. Eng. News 34*, 2812–2814, 2816 (1956).

PERRY J.W., and ALLEN KENT. *Documentation and Information Retrieval.* Interscience, New York, 1957, pp. 126–127.

STERN, JOSHUA. National Bureau of Standards instrument reference service. *Can. Chem. Processing 39*(10), 111–112 (1955).

TAUBE, M. Application of the uniterm System and the Association of Ideas to a Special Library File. U.S. Armed Forces Technical Information Agency document; AD 93969.

WARHEIT, I.A. Evaluation of library techniques for the control of research materials. *Am. Document. 7,* 267–275 (1956).

WELT, ISAAC D. Subject indexing in a restricted field. *Science 123,* 723–724 (1956).

WILDHACK, WILLIAM A., JOSHUA STERN, and JULIAN SMITH. Documentation in instrumentation. *Am. Document. 5,* 223–237 (1954).

Retrieval Questions from the Use of Linde's Indexing and Retrieval System

FRED R.WHALEY

The indexing of company reports at Linde's Tonawanda Laboratories is now in its fourth year (1–3). Retrieval questions began coming in at the start of 1955 and records have been kept since then. These questions have been analyzed both with respect to their quantitative distribution with time, and their qualitative distribution according to the logical type of question involved.[1]

FREQUENCY OF RETRIEVAL QUESTIONS

Inquiries are rather infrequent in the early stages of a non-conventional indexing system. The two principal reasons for this are: (1) the material in the index is rather sparse with a low probability of covering the work in question and (2) the system is strange to the technical men and they are unaccustomed to the new type of service it can provide. To realize maximum advantage of the system, experience is needed in asking questions of a depth or specificity previously impractical in a conventional system. Much information previously considered of no interest simply because it was difficult or impossible to locate can now be retrieved with speed and thoroughness (provided, of course, that it is part of the body of literature already indexed). It is not easy to instruct technical men in this new potentiality; they have to experience it.

The number of retrieval questions processed during the three years of operation is as follows: 1955, 38 questions; 1956, 65 questions; 1957, 130 questions.

This illustrates the slow start and the increase by geometric progression in the use of a non-conventional index that occurs in its first few years. The same experience was obtained by G.L.Peakes (4) at Bakelite Development Laboratories,

FRED R.WHALEY Technical Literature Coordinator, Tonawanda Research Laboratory, Linde Company, Division of Union Carbide Corporation, Tonawanda, New York.

[1] The paper by Saul Herner and Mary Herner, "Determining Requirements for Atomic Energy Information from Reference Questions," which has been placed in Area 1, also contains an analysis of reference questions.

another Division of Union Carbide Corporation. The material available in an index increases approximately by arithmetical progression, and if the content of an index were the only factor determining the use of the index, the latter should also increase arithmetically. The apparent discrepancy is largely due to the other most important factor, the overcoming of user inertia.

THE LOGICAL ANALYSIS OF RETRIEVAL QUESTIONS

Numerous workers in the field have pointed out the relationship between formal logic and the requirements of an information retrieval system (5–7,9,10). Most of the work has been on theoretical grounds with considerable disagreement regarding symbols to be used and the extent to which the problems of information retrieval are actually served by the concepts of formal logic. Rather than delve into the finer points of this problem, on which there are varying points of view, our purpose is to show that the logic of classes in its simplest and least sophisticated form is directly applicable and useful in Linde's retrieval system.

To serve this purpose best the logical symbols are defined pragmatically as they relate to actual units (cards) in the system itself. The relation they bear to the more rigorously defined symbols and operations in the literature will be evident.[2]

Linde's system is a collating system. Each concept used in indexing a report is assigned a term number, such as 4321. Let a capital letter, such as A, represent all the reports in the index which require term 4321. Then A represents the class of reports dealing with this term. Similarly B, Q, and X represent classes of reports dealing with three other respective terms. Any given class of reports so defined will be found readily as a deck in the file, since the term number is known and the cards are in order by term number. Each card identifies a report, so the entire class of reports, as well as each individual report, is identified by the deck.

A term number defines a primary class of documents and is symbolized by a capital letter in the logical expressions shown below. Certain logical operations can be performed on these primary classes, and the result of each operation is a new derived class of documents. For example, A and B can be matched by document number to see if any documents have these two term numbers in common (symbolized A·B). Before matching they first must be sorted in order by document number. During matching, the matched (and merged) cards are selected and non-matched cards are rejected. The selected cards (A·B) now symbolize a derived class of documents, which meet the logical

[2] The logical operations described herein result in identifying pertinent reports. As shown in (1) the final retrieval step is the visual identification of pertinent items within reports. This step is omitted from the present discussion, but it involves the same logical principles.

requirements of the expression, A·B. The multiplication sign means "in conjunction with," and the derived class (deck of cards in order by document number) can in turn be matched with the C deck to give (A·B)·C. Common sense and a little experience with the cards convinces one that the order in carrying out the steps does not affect the ultimate answer (the content of the final derived class), although order of operations may affect the time required to obtain the answer.

If an inquiry involves logical alternation, the plus sign indicates the operation, and it is translated "and/or." Thus if a question involves either class A or B, all the documents in both decks are required. The logical expression is A+B, and the corresponding operation on the cards is to combine both decks and sort them all together in order by document number. The combined deck is a new derived class of documents meeting the logical requirement, A+B. While conjunction can be carried out between only two decks in one operation, alternation can be carried out between any number of decks in one operation. In certain questions hundreds of decks (such as classes of chemicals meeting certain structural requirements) are combined into a single class by a single sorting operation to give (A+B+C+· · ·).

The operation of logical negation is symbolized by a minus sign and may be translated "without" or "with the exception of." This is used rarely; it will be explained more fully below in the discussion of the statistics.

The statistics are taken from the first 260 questions submitted to our information center. Of these, one question was impossible to handle, that is, our terminology had not been organized in a manner suitable for the question. Nine questions were of a type requiring two alternate approaches, and therefore were counted as two questions each. Thus the statistics contain 268 questions distributed into types as listed and illustrated in Table 1. Underlinings in examples illustrate the terms used. A double underline means more than one term represented. After each logical expression are shown the number of questions in that type and the percentage of the total.

Types 1 and 2, totaling 32.4%, are the only types not involving card matching. Answers are obtained here simply by selecting the proper decks from the ordered file and reading off the references shown on the cards. No machine work is required except to sort by document number for convenience.

Type 3 is the simplest of all conjunctions. All subsequent types may be expanded into combinations of simple conjunctions of the A·B type. For example, Type 5 may be expressed as A·B+A·C+· · ·. Types 5 through 9 are expressed in our system as combinations of conjunction and alternation. From the number of factors in the logical expression subtract one to get the number of machine matching operations required, i.e., one in Type 5, two in Type 9. If any one of these types is expressed in expanded form as a series of simple conjunctions, many more matchings are needed, sometimes numbered in the hundreds for a single question. The expanded expression for these types requiring many conjunctions is the only way they can be implemented by some retrieval systems, such as Taube's "Uniterm" (6) or Batten's "Peek-a-Boo" (8) systems. Much of the literature on these systems implies that most questions are of Types 3 or 4. Our experience shows that these types combined comprise only 21.4% of the total.

TABLE 1

Type	*Logical expression*	*Number*	*%*	*Example*
1	A	36	13.4	Information on γ-aminopropyl-trichlorosilane
2	A+B+· · ·	51	19.0	Information on any aminoalkyl-trichlorosilane
3	A·B	32	12.0	Refractive index of ferrocene
4	A·B·C· · ·	25	9.4	Cost estimates of equipment for engine testing
5	A·(B+C+· · ·)	56	21.0	Disproportionate of metallic subhalides
6	(A+B+· · ·) (Q+R+· · ·)	33	12.3	Any thermodynamic properties of ferrocene or its derivatives
7	A·B· · ·(Q+R+· · ·)	23	8.6	Heat of formation of any one of a group of chemicals
8	A(B+C+· · ·) (Q+R+· · ·)	6	2.2	Viscosity as a function of temperature or shear in dimethylsilicone oils
9	(A+B+· · ·) (Q+R+· · ·) (X +Y+· · ·)	2	0.7	Any of various clay minerals compared directly with any zeolites for either adsorptive or catalytic uses
10	A–A·B	2	0.7	Preparation of a chemical not using a Grignard reagent
11	A·B·A·B·C	2	0.7	The reaction of two chemicals where a third chemical is definitely not formed

The most frequent single type is Type 5 (21%) where a particular term is essential to the answer and a conjunction with any one of several other terms is required. Types 5 through 9, which involve alternations as well as conjunction, comprise 44.8% of the questions.

Types 10 and 11 involve the minus sign (negation) and are usually expressed differently in the literature:

Type	*Our expression*	*Usual expression*
10	A–A·B	$A - B$ or $A \cdot \overline{B}$ or $A(1 - B)$[a]
11	A·B–A·B·C	$A \cdot B - C$ or $A \cdot B \cdot \overline{C}$ or $A \cdot B(1 - C)$[a]

[a] Unity (1) is the symbol representing all the documents in the indexed file.

Our way of expressing negation is directly related to the way we handle the cards. Negation has meaning in relation to a pack of cards only in that some of the cards are rejected because they represent a class of document not wanted. Thus for Type 10, A—A·B, we select deck A from the file and reject from it the part that matches deck B. The expression, A—B, has no meaning to us since we cannot tell from decks A or B alone which cards to reject without a matching operation.

The expression, $A \cdot \overline{B}$, or A(1–B), means the conjunction of A with "not B." If A were matched with the entire remainder of the file excluding B (about 200,000 cards), this operation would reject from the A deck only those documents indexed with the A and B terms alone, which is an inadequate answer as well as a very impractical procedure. Of the various expressions for this type of question, which are logically equivalent, only the one we use is operative for our system. For Type 11, we find the matches between decks A and B and reject from this pack of cards the matches between it and deck C. Again the usual expressions are inoperative.

We make use of a role code appended to our term number to mean that the item emphasizes the absence of a particular term. This device cuts down on the frequency of logical negation in analyzing questions. As the examples show, there are some situations where the negative role will not serve, and logical negation must be used (1.4%).

More sophisticated systems will require refinements in the logic employed. For example, our system does not allow distinction based on order of terms, (A·B distinguished from B·A), although we accomplish somewhat the same end by appending role numbers to the term number to show a particular context, such as object of a chemical preparation. We have not developed a good means of bracketing terms in the indexing step. For example, an item dealing with an <u>aluminum flange</u> on a <u>copper tube</u> (terms underlined) might be retrieved falsely by some one looking for aluminum tubing, unless aluminum and flange (as well as copper and tube) are precoordinated to give, in the indexing step, [A·B] · [C·D]. This document would be retrieved by a question on an aluminum flange in equipment involving copper, [A·B]·C, or copper tubing in some equipment involving aluminum, A·[C·D], but not by a question asking for aluminum tubing, [A·D], with or without copper equipment also involved, [A·D]·C. These and other refinements must be investi gated for any solution to the growing problem of information retrieval in the world's technical literature.

DESIRABILITY OF DISCUSSING QUESTIONS WITH THE INQUIRER

G.L.Peakes (4) and others have pointed out that retrieval questions as originally received may not truly express what the questioner actually wants. For example, a question on "thermal conductivity of aluminum alloys" was found, on discussing it with the inquirer, to mean "overall heat transfer of fins made from aluminum alloys." This led to using more appropriate terms to arrive at the desired references. This give and take concerning a question before processing it is frequently called "negotiating" a question.

An inquirer may want a quick typical answer rather than a thorough and complete one. For example, his question may be expressed logically as A·B·C·D, which is quite specific. If an answer or two is obtained he will be satisfied. On the other hand, he may instruct us that he wants everything that might possibly have a bearing on his question, even where the authors did not recognize all four terms in conjunction. Consequently, we will give him the answers from A·B·C·D as the best answers (most likely to pertain to his question), but will also include as possible answers the remaining conjunctions A·B·C and A·B·D, assuming A and B to be essential terms in this example. The answers obtained from the more general treatment will contain more extraneous material, but this is the price paid for increased thoroughness in any conventional or non-conventional retrieval system yet designed.

SUMMARY OF LOGICAL ANALYSIS

The logic required to guide us in the handling of cards for information retrieval is very simple. Each class of reports (capital letter) involving a certain concept means to us the deck of cards having the term number assigned to that concept, with each card identifying a particular document. If two or more classes are connected by plus signs, the decks are sorted together by document number and treated as a new class in subsequent operations. If two classes are connected by a multiplication sign, they must be matched, and the matched cards comprise a new class to be used in subsequent operations, if necessary. If two classes are connected by a minus sign, the class following the minus sign must be a part of the class preceding the minus sign. The minus sign simply means rejecting from a deck of cards that portion which matches another deck.

Retrieval questions on a single term (Type 1) involving neither conjunction, alternation, nor negation comprise 13.4% of the total. Conjunction alone

(Types 3 and 4) involves 21.4%. Alternation alone (Type 2) accounts for 19%. Combination of conjunction and alternation (Types 5 through 9) comprises 44.8% and leaves only 1.4% for questions involving negation (Types 10 and 11).

We do not employ some of the refinements in logical analysis required for the major problem of information retrieval in the world's technical literature. For a medium-sized document center such as ours dealing primarily with company reports and limited by relatively inexpensive machinery, the logic we employ appears to be quite adequate.

REFERENCES

1. FRED R.WHALEY. A deep index for internal technical reports in *Information Systems in Documentation,* edited by J.H.Shera, A.Kent, and J.W.Perry. Interscience Publishers, New York, 1957.
2. FRED R.WHALEY. A deep index for internal technical reports in *Multiple Aspect Searching for Information Retrieval,* edited and published by the Armed Services Technical Information Agency, Washington, D.C., 1957.
3 FRED R.WHALEY. Linde Company (System) in *Non-Conventional Technical Information Systems in Current Use,* edited and published by the Office of Scientific Information, National Science Foundation, Washington, D.C., 1958.
4. G.L.PEAKES. Experience with the unit card system for report indexing in *Information Systems in Documentation,* edited by J.H.Shera, A.Kent, and J.W. Perry. Interscience Publishers, New York, 1957.
5. J.W.PERRY, A.KENT, and M.M.BERRY. *Machine Literature Searching,* Interscience Publishers, New York, 1956.
6. MORTIMER TAUBE and ASSOCIATES. *Studies in Coordinate Indexing,* Documentation Incorporated, Washington, D.C., 1953. Also, “The Distinction Between the Logic of Computers and the Logic of Storage and Retrieval Devices,” Revised Edition (AFOSR TN 57–165), September, 1957.
7. V.P.CHERENIN. *Nekotoryye problemy dokumentatsii i mekhanizatsiya informatsionnykh poiskov* (Certain Problems of Documentation and Mechanization of Information Search), Moscow, U.S.S.R., 1955.
8. W.E.BATTEN, Specialized files for patent searching in *Punched Cards,* edited by R.S.Casey and J.W.Perry. Reinhold, New York, 1951.
9. J.W.KUIPERS, A.W.TYLER, and W.L.MYERS, A Minicard system for documentary information in *Information Systems in Documentation,* edited by J.H. Shera, A.Kent, and J.W.Perry. Interscience Publishers, New York, 1957.
10. DON D.ANDREWS. Interrelated Logic Accumulating Scanner (ILAS), Patent Office Research and Development Reports, No. 6. June 25, 1957.

Classification with Peek-a-boo for Indexing Documents on Aerodynamics: An Experiment in Retrieval

R.C.WRIGHT and C.W.J.WILSON

ABSTRACT. A new classification and peek-a-boo cards were used for the analytical indexing of documents on aerodynamics and 100 test questions were put to the system. Degree of success in retrieval is assessed, input and search times are measured, and causes of failure are examined. The combination appears to offer a promising retrieval system for complex but well-defined subjects.

Towards the end of 1956, it was proposed to prepare and (as far as possible within security limits) to distribute to industry a subject index to the reports and technical notes issued by the Aerodynamics Department of the Royal Aircraft Establishment. This index, which might later be extended to include other series of reports on aerodynamic subjects, would be based on a suitable classification system permitting conjunctive specification, and would use a clerical device of the peek-a-boo type (1). The system would employ a minimum of expensive equipment and, if successful, might be applicable to other fields requiring detailed analytical indexing of a comparatively small collection. First, however, it would be necessary to select a suitable classification of aerodynamics, do some trial indexing, and carry out tests on the system. The purpose of this paper is to describe the selected classification, the posting and searching operations, and the tests carried out.

PURPOSE OF THE EXPERIMENT

Main reasons for undertaking the project were:

1. A reliable analytical index to the 6000 reports in the series was necessary to the Establishment and would no doubt be welcomed by the aircraft industry.
2. Mr. J.Seymour, a former Librarian of the Aerodynamics Department, had recently developed a classification for aerodynamics which appeared to

R.C.WRIGHT and C.W.J.WILSON Royal Aircraft Establishment, Farnborough, England.

have many advantages for use in a retrieval system. Experience in its practical application was desired.

3. Tests had already been carried out with a Uniterm system for cataloguing aircraft structural data (2), and on the N.L.L. card catalogue of aerodynamic measurements (3, 4). Similar evaluation of a peek-a-boo device based on a suitable subject classification was considered desirable.
4. It was hoped to ascertain whether an analytical indexing and retrieval project in the complex subject of aerodynamics could be designed and operated by professional librarians without specialist subject knowledge, to the satisfaction of the aerodynamicists requiring the information.

PHYSICAL FORM OF THE INDEX

In its final form, the index would be made up of (1) a register, i.e., a list of all documents indexed by the system, in register serial number order; (2) a manual, consisting of *(a)* an introduction to the system, *(b)* a list of subject headings in classified order, *(c)* an alphabetical index to the subject headings; (3) index sheets—one for each subject heading and for each designated aircraft and aerofoil; register serial numbers of documents to be indicated by the positions of punched holes.

For testing purposes, however, it was decided to use only the classification schedule itself and standard 80-column Hollerith cards. Special cards or plates having a higher capacity might later be desirable, but their design, and that of suitable punching and viewing devices, could be left until the retrieval capabilities of the system had been investigated.

CHOICE OF CLASSIFICATION

After consideration of the known enumerative classifications of aerodynamics, including U.D.C., N.A.C.A., and the N.L.L. (all of which are used to some extent in the Establishment), it was decided to adopt the classification system devised by Mr. Seymour with this project in mind. After comments had been sought from subject experts, the classification was adopted with slight modifications; it appears as Table 2, together with the number of times each code was used in indexing the first 700 documents. It was expected that further modifications would follow an initial trial period of indexing.

INDEXING

THE INDEXERS

The six indexers were professional librarians employed in the Main and Departmental libraries of the Establishment. One of them (A) was librarian of the Aerodynamics Department: the remainder (B-F) had experience in aeronautical libraries but no special knowledge of aerodynamics. About half of the indexing and searching was to be done by A, so that the effect of familiarity with the subject and with the schedules might be assessed.

Before starting to index, the indexers met to discuss the schedules and to resolve as far as possible differences of opinion as to the meanings or the use of particular headings. Each person then indexed a number of test pieces, and variations were discussed. It was intended to introduce further test pieces at later stages in order to see whether improved correlation came with experience of the system.

PROCEDURE

For test indexing the selection of documents was not limited to R.A.E. documents, some Institute of Aeronautical Sciences preprints and National Advisory Committee for Aeronautics Research Memoranda being included in order to achieve wider subject coverage, particularly of recent material.

The intention was to code all those features of a document which might conceivably be used later as part of an information request. A detailed study of each document was therefore necessary, particular attention being given to the summary, conclusions and illustrations. Postings were made on a 5 by 3-inch card ruled in ten columns as follows:

0	1	2	3	4	5	6	7	8	9	New Codes
										Serial No.

Thus, to indicate that a document dealt with a fighter aircraft, the code 157 taken from the schedules would be entered as 15 in column 7. All relevant codes were entered, so that the issue date 1953 would be coded as 103, 104, 105, 106, and Mach number 1.3 would be 121, 122, 123, 124. Whenever a sub-heading was used, the main heading above it was also coded. For peek-a-boo use, these codes were subsequently transferred to lists in code number order and punched on Hollerith cards.

The only materials and equipment used were common stationery items and a Hollerith hand punch. The processes involved were:

(a) Examination of the document and noting appropriate codes on the 5`×3` card illustrated above.

(b) Listing the serial numbers appropriate to each code number.

(c) Hand punching the Hollerith cards, which already have typed on them the code numbers, by which they are arranged.

Times taken for these operations are discussed later. The most time consuming, and that needing the highest-paid staff, was (a). Most documents were assigned about 30 codes. At first, indexers were taking as long as 30 minutes for each document: practice, familiarity with the schedules and increasing knowledge of the subject soon reduced this to 15–20 minutes per document. It is doubtful whether an average of about 15 minutes can be much reduced, in view of the careful study of each document which is necessary to any system using highly analytical indexing. The average salary of the indexers employed worked out at 10/– ($1.40) an hour.

DESCRIPTION OF TESTS

The test programme is in two parts:

(a) When about 750 documents have been indexed. To ascertain the effectiveness of the classification, to make any necessary modifications to the schedule, and to remove causes of discrepancy between indexers. Also to assess the cost of indexing and posting and the speed of searching operations.

(b) When about 3000 documents have been indexed. To ascertain the number and relevance of documents retrieved in response to a given question. Also, on the basis of search times, to ascertain the scope and form of the final index *(a)* for internal use and *(b)* for possible issue to other aeronautical libraries. Only part *(a)* of the programme is discussed in this paper.

When 700 reports had been indexed (500 R.A.E. documents and 200 of U.S.A. origin), ten R.A.E. aerodynamicists were asked to collaborate by selecting 10 documents each from the collection and framing one question on each document. The questioners had not seen the classification schedules, and were asked to frame in their own words the sort of question that might occur in their daily work. The questions appear in Table 1. Searching was divided about equally between indexers A and B. Success was defined as the retrieval of the document on which the question had been based. Obviously this had to be limited in some way, since a search which retrieved the subject document together with about 50% of the collection could hardly be regarded as an unqualified success. Such a case could be due to faults in retrieval; it could also be due to the question having been so general that it was bound to produce a large number of answers.

After discussion, an arbitrary limit of 5% of the collection (35 documents) was set as the maximum for a search to be regarded as successful. Searches producing more than this number have been classed as "partially successful" provided that the subject document is included.

When a search was unsuccessful, a second search was usually made, either by the same searcher or by the other. Sometimes this consisted simply of widening the original search by discarding one or more of the factors forming the specification. In other cases completely new headings were selected to form a new search plan.

Usually not more than two searches were made, but there were a few cases in which a third search was considered advisable.

ANALYSIS AND DISCUSSION OF RESULTS

PERCENTAGE SUCCESS

Of the 100 questions put, 54 were answered at the first search, a further 20 by a second search, and a further 4 by a third. In 7 cases the search was "partly" successful (subject document retrieved, but with more than 5% of the collection), and in 15 cases the search failed completely. An overall success percentage of 78 (plus a further 7% of cases in which the field was appreciably narrowed) is not unsatisfactory having regard to the circumstances in which the test was made. Many of the questions would, in normal information work, have been referred back for amplification.

CAUSES OF FAILURE

(a) Causes of complete failure

	Question numbers														
	1	7	12	32	34	36	39	40	51	53	57	66	81	88	93
Misleading question					X				X						
Question misinterpreted	X	X													
Indexing omission	X		X		X	X	X	X		X	X	X	X	X	X
Schedule inadequacy	X			X								X			

(b) Causes of partial success

	Question numbers						
	17	23	24	26	28	61	70
Question insufficiently specific			X		X		
Indexing omission	X			X	X	X	X
Schedule inadequacy		X					

(c) Causes of earlier failure when third search successful

	Question numbers 37	41	48	67
Question misinterpreted				X
Indexing omission	X	X	X	

(d) Causes of earlier failure when second search successful

	Question numbers 4	5	11	22	25	31	33	35	49	55
Misleading question					X					
Indexing omission	X		X			X		X	X	X
Faulty search			X	X						
Search too specific					X			X		
Schedule inadequacy		X					X			

	Question numbers 56	60	65	72	76	84	86	87	96	97
Indexing omission	X	X	X		X	X	X	X		X
Search too specific	X			X						
Clerical error in posting		X								

(e) Summary. A total of all the above single or contributory causes of complete or partial failure may be helpful in pinpointing the chief weaknesses of the system.

Semantics		
Misleading question	3	
Question insufficiently specific	2	
Question misinterpreted	3	8
Input stage		
Indexing omission	34	
Schedule inadequacy	6	
Clerical error	1	41
Output stage		
Faulty search	2	
Search too specific	4	6
Total		55

It will be noted that the largest single factor was indexing omission, and this in turn is believed to be attributable mainly to lack of subject knowledge. Next largest is the "semantics" group, failure in communication between the questioner and the searcher.

The six cases of schedule inadequacy, together with other omissions noted in indexing the first 700 documents, will lead to minor revision of the classification. At this time, reconsideration will be given to the necessity for retaining some headings which have hardly been used at all, and others which have been used so often as to render them almost useless for retrieval purposes.

The six cases of failure at the output stage are not considered significant: none persisted beyond a first search, and with peek-a-boo it is possible to make several searches in a few minutes.

NUMBER OF CODES USED IN INDEXING

Average number of codes used in indexing 700 documents	28.96
Average number of codes used in indexing the 54 documents successfully retrieved at the first search	30.59
Average number of codes used in indexing the 15 documents which could not be retrieved	30.07

Indexer A used an average of 30.50 codes for each report.
Indexer B used an average of 32.94 codes for each report.
Indexers C`F used an average of 29.12 codes for each report.

All these averages, except the first figure of 28.96, are based only on the 100 documents on which questions were asked.

ANALYSIS BY INDEXER

Indexer	*Questions*	*Documents indexed*	
A			
Successful first search	9, 10, 14, 18, 19, 20, 21, 27, 29, 42, 44, 45, 46, 47, 50, 52, 62, 63, 64, 71, 75, 79, 80, 82, 83, 85, 89, 90, 91, 95, 98, 99, 100	33 (55%)	49 (81.6%)
Successful second search	4, 5, 11, 22, 25, 31, 33, 35, 55, 65, 72, 84, 86, 97	14 (23.3%)	
Successful third search	48, 67	2 (3.3%)	
Partly successful	28, 61	2 (3.3%)	
Failed	7, 32, 39, 40, 51, 53, 66, 81, 88	9 (15%)	
Total		60	
B			
Successful first search	3, 6, 38, 59, 68, 73, 74, 92, 94	9 (56.2%)	12 (75%)
Successful second search	60, 76, 96	3 (18.7%)	
Partly successful	17, 24, 70	3 (18.7%)	
Failed	1	1 (6.2%)	
Total		16	
C, D, E, and F			
Successful first search	2, 8, 13, 15, 16, 30, 43, 54, 58, 69, 77, 78	12 (50%)	17 (70.8%)
Successful second search	49, 56, 87	3 (12.5%)	
Successful third search	37, 41	2 (8. 3%)	
Partly successful	23, 26	2 (8.3%)	
Failed	12, 34, 36, 57, 93	5 (20.8%)	
Total		24	

All indexers used about the same average number of codes per document, but the success achieved varied with the indexer's subject knowledge. Indexer A was in charge of a specialised library for aerodynamicists; B, in charge of a more general aeronautical library, had considerable experience in aviation and in library services; C to F had served in aeronautical libraries for periods varying from 5 years to only a few months.

RELEVANCE OF DOCUMENTS RETRIEVED

In the N.L.L. tests (3, 4) searchers were able to make a rough assessment of the relevance of each document retrieved by an examination of the detailed information printed on the N.L.L. punched cards. In the present case it would have been necessary to refer each document, or an abstract of it, to the questioner. This being a time-consuming operation, it was decided to do so only on a sample of 20 successfully answered questions. These were further restricted to cases in which 3 to 12 documents had been thrown up in the retrieval process. In 20 cases, therefore, questioners were presented with abstract cards or documents and they decided whether the documents retrieved, in addition to the ones on which their questions were based, were truly relevant in terms of their questions as set.

Total documents retrieved for the 20 questions came to 146 (the number of aspect cards correlated varied from 1 to 5, but was usually between 2 and 4). It was found that 70 documents were relevant, 18 were of marginal relevance, and 58 were irrelevant. This makes a broad ratio of 3.5:1:3, which appears to be satisfactory, although it is clearly desirable to reduce the marginally relevant figure. This can probably be achieved by improved indexing.

TIME STUDIES

INPUT

Accurate time studies were made of three main aspects of the input stage of this project. These were: times taken to index each document, times taken to post code numbers to a ledger, and times taken to transfer this information to punched Hollerith cards.

(a) *Time to index documents.* Indexer A, who was responsible for half of all indexing, timed himself on 17 occasions. This was done only after considerable experience of indexing difficulties and detailed knowledge of the schedules had been gained. 153 documents were indexed in 31 hours 5 minutes, which gives 13.5 minutes per document.

(b) *Time to post index numbers.* All papers indexed had their codes recorded on the card described in “Procedure” under “Indexing.” These codes were

then transferred to a plain ledger to facilitate eventual punching. In 18 operations, codes for 474 documents were posted by 2 persons in 19 hours 16 minutes, which gives 2.4 minutes per document.

(c) *Time to punch Hollerith cards.* Codes for 500 documents were transferred from the ledger and punched on Hollerith cards in 19 hours 30 minutes. This was done in 12 operations by 2 persons, at an average of 2.34 minutes per document.

Total times for each document subject-indexed by the system described therefore consist of 13.5 minutes for indexing, 2.4×2 minutes for posting, and 2.34×2 minutes for punching, or 22.98 minutes. These are the three time-expensive parts of the programme, but, if an allowance of 5 minutes for extracting or filing papers or associated clerical processes is made, it is reasonable to assess inclusive time for subject indexing and all allied clerical effort as an average of 28 minutes per document. Half of this is relatively expensive indexing. No great reduction in any of these times can be expected.

OUTPUT

The output stage of this system consists firstly in relating a question to the classification schedules, in carefully assessing the best aspect cards to be used,[1] in correlating these cards and reading the holes punched all through. With practice, these were found to be rapid processes which, depending on the complexity of the question, could be performed within 2–10 minutes. The second phase consists of removing the relevant serially numbered abstract cards from the file, and offering them to the questioner, who may then wish to see some of the actual documents represented by the cards. The total output stage should not normally exceed 15–20 minutes.

CONCLUSIONS

The combination of subject classification and peek-a-boo appears to have many advantages as a means for the analytical indexing of documents in a well-defined subject such as aerodynamics. There would appear to be a market for the commercial development of “miniaturised” cards or plates having a capacity of 20,000 or more positions and of appropriate punching and viewing equipment.

[1] Bernier (5) has found, theoretically, that “discovery of unexpected documents by use of a manipulative, correlative index is usually highly improbable if the number of terms taken at a time in searching is four or more….” It was found, as described in the paragraph on relevance of documents retrieved, that the number of peek-a-boo cards correlated usually fell within the range 2–5.

Best results have been obtained by indexers with subject knowledge, and "production" indexing should be undertaken only by such indexers.

The classification used was basically sound, but requires expansion in some areas and contraction in rather more, so as to achieve less variation in the number of times code-numbers are used. In many cases scope notes are required.

A survey (6) of the report literature actually in use by a typical group of aerodynamicists shows that 98% of the reports were less than 10 years old and 83% less than 5 years old. This finding is likely to influence schedule revision and may necessitate re-appraisal of the scope of the present project. It also tends to minimise one criticism of peek-a-boo, i.e., that it is necessary at intervals, depending on card capacity and accession rate, to start a completely new index.

REFERENCES

1. WILDHACK, W.A., STERN, J., SMITH, J., Documentation in instrumentation. *American Documentation 5,* 223–237 (1954).
2. CLEVERDON, C.W., THORNE, R.G., A brief experiment with the Uniterm system of co-ordinate indexing for the cataloguing of structural data (unpublished).
3. VESSEY, H.F., Test of N.L.L. card catalogue of aerodynamic measurements (unpublished).
4. VESSEY, H.F., SEYMOUR, J.R., Test of N.L.L. card catalogue of aerodynamic measurements. II (unpublished).
5. BERNIER, C.L., Correlative indexes and the blank sort. *American Documentation 9,* 32–41 (1958).
6. WILSON, C.W.J., Report literature used by aerodynamicists (unpublished).

TABLE 1 *The 100 questions and the searches based on them*

Question No.	Question	Search plan	Documents retrieved	Indexer	Codes used	Searcher	Success or failure	Remarks
1	Effect of incidence on pressures recorded on a Hunter aircraft.	(*a*) 365 362 (Hunter)	(*a*) 14	B	21	B	F	Peek-a-boo cards for aircraft names are not yet available, but would normally be used. In this case neither "Hunter" nor "365-Incidence" was indexed.
		(*b*) 365 417 362	(*b*) 1			A	F	Searcher B misunderstood the question as relating to aerodynamic pressures: document related to pitot-static tube pressures. "Pitot static tubes" had been noted by indexer as a desirable new heading.
2	Aerodynamics of struts	349 264	10	C	36	A	S	
3	Spinning tunnel instrumentation	397 417	1	B	22	A	S	
4	Early methods of measuring rates of climb (relative merits of barograph and cinematograph)	(*a*) 417 369 378	(*a*) 0	A	26	B	F	"417-Instruments" not indexed.
		(*b*) 369 378	(*b*) 14			B	S	
5	Applications of the Ludwig-Tillman skin friction formula	(*a*) 128,137 136,154 349,352	(*a*) 4	A	28	A	F	Indexers had noted skin friction as a necessary new indexing term.
		(*b*) 113,128 137,352	(*b*) 12			A	S	
6	Effect of rate of approach to a stall on the stall C_L	365 350	26	B	47	B	S	378, if used, would reduce the number of documents to 3.

Question No.	*Question*	*Search plan*	*Documents retrieved*	*Indexer*	*Codes used*	*Searcher*	*Success or failure*	*Remarks*
7	Transient analysis techniques of frequency response measurement using pulse type inputs (assessment of use)	(*a*) 373	(*a*) 48	A	23	A	F	Subject document discussed dynamics of helicopter rotor control systems.
		(*b*) 417 391 416	(*b*) 2			B	F	
8	Correlation of theory and experiment pressure distributions around swept wings at high subsonic Mach numbers	362 206 125 123	20	C	26	B	S	
9	Effects of wing dihedral on lateral stability	195 232 355 357	3	A	35	A	S	
10	Creep of turbine blades	327 386	4	A	18	B	S	
11	How can the exhaust fumes from an aero-engine be dispersed when doing full-scale tests in a closed circuit wind tunnel?	(*a*) 326 339 392 391	3	A	29	B	F	"377-Ventilation" might well have formed part of the search plan.
		(*b*) 339 392,391	13			B	S	"326-Engine" not indexed.
12	What is the effect of endplates (tip fins) on lift slope and pitching moment of a straight wing at high subsonic Mach numbers?	(*a*) 112,123 195,196 205,302 320,235 132	0	D	30	A	F	Indexing omissions. Not indexed under end-plates, wing, monoplane, or interest-at-wing-tip.
		(*b*) 320,358 123,125	6			A	F	

13	What is the effect of movement of C.G. position on spinning characteristics of aircraft of conventional layout and unswept wings?	193 195 205 251	6	F	24	A	S	
14	What will be the effect of the tip vortices from the main wings on the tail surfaces of a cruciform-winged missile with indexed cruciform controls?	143 198 164	5	A	37	B	S	
15	At sonic or supersonic speeds, what is the load distribution on a cropped delta wing with a specified twist distribution across the span?	122 195 203 363	14	C	26	A	S	
16	What is the effect of surface roughness on the drag of an aircraft?	341 352 339 193	2	F	28	B	S	
17	What information exists on methods of reducing the landing run of high speed aircraft?	(*a*) 193 367 377	(*a*) 15	B	34	A	F	
		(*b*) 193 367	(*b*) 38			A	P	"377-Operation/Design" not coded, but its division "378-Performance" was coded. Indexing omission.
18	How can aerodynamic derivatives due to yawing motion be measured in a wind tunnel?	359 392 417	4	A	35	B	S	
19	Does the nature of the surface terrain significantly affect the pattern of gusts encountered by an aircraft?	153 377	4	A	24	A	S	

Question No.	*Question*	*Search plan*	*Documents retrieved*	*Indexer*	*Codes used*	*Searcher*	*Success or failure*	*Remarks*
20	What is the best speed to fly a sailplane in soaring conditions for various lift/drag ratios and thermal strengths?	178 419	1	A	13	B	S	
21	Theoretical analysis of cranked wings.	113 195 204	3	A	27	B	S	
22	Hot wire amplifiers.	(*a*) 388 411 417	3	A	19	A	F	
		(*b*) 388 417	8			A	S	
23	Wind tunnel facilities at Arnold Engineering Development Centre, Tullahoma.	391 392	74	C	35	A	P	Use of "112-Experiment" reduces to 66 documents, but this type of question is better answered by use of conventional index. The code "115-USA" was to be used only for documents originating in the U.S. but might have been appropriate in this case.
24	Pressure distribution on fuselage at supersonic speeds.	362 240 122	42	B	28	B	P	Insufficiently specific question to narrow the search further. 6% of the lection may well be relevant to the question.
25	Calculation of biplane lift at subsonic speeds.	(*a*) 113 197 350 125	0	A	28	B	F	

		(*b*) 113 197	2			B	S	The subject document actually discusses possibilities of a biplane for supersonic flight.
26	Separated flow: theoretical treatment.	(*a*) 113 142	(*a*) 12	C	17	B	F	"113-Theory" not coded.
		(*b*) 113 142 128	(*b*) 12			A	F	"142-Separation" alone is successful, but produces 36 documents.
		(*c*) 142	(*c*) 36			B	P	
27	Wind tunnel tests on X-1E aircraft.	(*a*) Aircraft names		A	32	A	S	(*a*) Aircraft names index has not yet been transferred to peek-a-boo, but would obviously be successful.
		(*b*) 112 115 390 392	(*b*) 5			A	F	(*b*) Failure due to clerical error. 392 was indexed but not punched.
		(*c*) 112 115 390	(*c*) 12			A	S	
28	Shock wave interaction.	(*a*) 148 364	(*a*) 8	A	18	B	F	
		(*b*) 148	76			B	P	"364-Interference/relative position" not indexed.
29	Downwash behind delta wing. Subsonic speeds.	125 144 203	11	A	42	A	S	
30	Mechanical design of flexible walls for wind tunnels.	408 391 387 392	1	C	35	B	S	

Question No.	Question	Search plan	Documents retrieved	Indexer	Codes used	Searcher	Success or failure	Remarks
31	What data are available on gust loads and their alleviation?	(*a*) 153 363 419	2	A	12	B	F	
		(*b*) 153 419	4			B	S	"363-Loads" not indexed.
32	Data required on the drag of struts, effect of section shape, fineness ratio, etc.	(*a*) 264 352	4	A	37	A	F	"264-Strut mounted" not indexed, and the schedules do not provide for struts themselves.
		(*b*) 227 247 or 248 352	0			B	F	
33	What information is available on the drag penalty of various radiator schemes?	(*a*) 352 155	6	A	28	B	F	Radiator cannot be expressed in the existing schedules.
		(*b*) 352 326	15			B	S	
34	Are there any flight data on the measurement of control panel deflections?	(*a*) 405 302 363	0	F	23	B	F	The subject document describes only wind tunnel experiments.
		(*b*) 406 302 363	1			B	F	"302-control device" not indexed.
35	What flight experience is available on the effect of drop tanks (strut or tip) on performance?	(*a*) 112, 259 261, 274 339, 405 378	(*a*) 0	A	47	A	F	
		(*b*) 112 259, 261	(*b*) 5			A	S	

36	How does wing tip shape affect the wave drag (tunnel data required).	(*a*) 392 235 352 222	7	E	41	B	F	"235-tip" not indexed.
		(*b*) 112, 195 235, 352	12			A	F	
37	Has any flight work been done with an aircraft fitted with a castoring undercarriage?	(*a*) 280	9	F	32	A	F	"280-Undercarriage" not indexed.
		(*b*) 112, 139, 367	2			A	F	
		(*c*) 112, 405, 367	14			A	S	
38	General information on seaplane tank techniques.	402	12	B	24	B	S	No co-relation necessary. Conventional index would serve just as well.
39	What is the effect of body cross-section shape on the isolated body moments in pitch and yaw?	(*a*) 240, 257 354, 358 359	1	A	40	A	F	The combination 240, 354 would have been successful, but with 39 other documents retrieved.
		(*b*) 358, 359, 354 240	10			B	F	"359-directional" and "257-section" not indexed.
40	What information is there on the use of spoilers for providing roll control at transonic and supersonic speeds?	(*a*) 319 356 357 122	3	A	25	B	F	"319-spoilers" and "356-control" not indexed.
41	The variation in hinge moment at low speeds due to flap angle, for a simple split flap near the half semi span position and close to the trailing edge. Wind tunnel results.	(*a*) 112 126, 317 354, 360	1	F	31	A	F	
		(*b*) 392, 317 354, 360	1			B	F	
		(*c*) 317, 392	12			B	S	Hinge moments (354,360) not indexed.

Question No.	*Question*	*Search plan*	*Documents retrieved*	*Indexer*	*Codes used*	*Searcher*	*Success or failure*	*Remarks*
42	Number of fatal accidents caused by fire in an aircraft during flight on American domestic airlines.	159 382	6	A	11	A	S	
43	Method for deducing suitable cross-sectional area distributions for a narrow delta wing (unswept trailing edge) for low zero-lift wave drag.	203 148	6	E	24	B	S	"229-thickness" would further reduce to 3 documents.
44	The jettison characteristics at low speed of a Blue Jay mounted on pylons under the wing of an aircraft.	371 164	2	A	36	B	S	Peek-a-boo cards representing proper names of aircraft, missiles, etc., have not yet been prepared, but would be used in practice in this case.
45	Longitudinal dynamic stability derivatives of a simple infra-red homing missile configuration, between $M = 0.7$ to 1.4.	123, 124 164, 355 358	5	A	45	A	S	
46	Theoretical estimation of wave drag at supersonic speeds of a slender body of elliptic cross-section with discontinuities in profile slope.	352 122 240 254	6	A	29	B	S	
47	The calculation of spanwise loading for an unswept uncambered oscillating wing by a "lifting line" technique.	(*a*) 195 205, 363	25	A	24	A	S	
48	A method for estimating the	(*a*) 363, 386	10	A	12	A	F	

	buckling load of flat sandwich panels, with the loading edges rigidly clamped.	(*b*) 109, 386	3			A	F	"363-Loading" not indexed.
		(*c*) 113, 386	17			A	S	
49	An assessment of the jet flap aerofoil as a practical means of control for an aircraft.	(*a*) 145 302, 318	3	C	38	A	F	
		(*b*) 325	10			A	S	"145-Jets" and "318-Other flap" not indexed.
50	A method of reducing heat transfer to blunt bodies by air injection.	154 240 243 347	1	A	31	B	S	
51	How can the thrust and torque of a windmill be calculated?	327 353	3	A	17	A	F	Paper is on autogyros. "327-Windmill" not indexed.
52	Experimental data on variation of heat transfer coefficient round the front of a hemisphere at high Mach numbers.	112, 120 154, 240 243, 255	2	A	28	A	S	
53	What methods are available for estimating the effect of incidence on the yawing moment due to sideslip at high supersonic speeds?	(*a*) 120 354 359 366	4	A	37	B	F	"354-moments" and 366-sideslip" not indexed.
		(*b*) 122 354, 359 365, 366	0			A	F	
54	Is there any information on the conditions leading to "intake buzz?"	285 136 148	11	F	27	B	S	
55	Methods for calculating the total pressure loss behind an aerofoil at low speeds.	(*a*) 109 126, 195	14	A	22	A	F	"109-Method for calculation" not indexed.
		(*b*) 126 195, 144	25			A	S	

Question No.	Question	Search plan	Documents retrieved	Indexer	Codes used	Searcher	Success or failure	Remarks
56	Comparisons between flight and wind tunnel tests of the lift increment due to flaps with blowing.	(a) 112, 305 306, 315	2	F	45	A	F	
		(b) 112 146, 315	2			A	S	
57	What methods are available for calculating the loading due to aileron deflection on a swept wing?	(a) 109 206 363	5	F	20	A	F	"206-swept" not indexed. Subject document did not relate specifically to swept wings.
		(b) 206, 363	35			B	F	"363-loading" not indexed.
58	What machine aids were used in World War II for stability and control calculations?	355 356 389	5	C	20	B	S	
59	List available design studies of supersonic aircraft.	122 193, 377	12	B	45	A	S	
60	What information is available on the normal force/incidence curve of bodies of revolution at incidences above 20° and speeds near Mach 2?	(a) 253 121, 124 365	0	B	40	B	F	"365-High incidence" not indexed.
		(b) 253 121, 124 392	14			B	S	
61	The effectiveness of different types of roughness in thickening turbulent boundary layers at various free-stream Mach numbers.	(a) 137 140, 341	5	A	24	A	F	
		(b) 137, 140	38			A	P	"341-Surface condition" not indexed.
62	The effects of the temperature and quantity of bleed flow on base pressure at supersonic speed.	122, 240 256 362	11	A	46	A	S	

63	The effects of side-support struts on the base pressure of a model at supersonic speeds.	392 122 362, 240 256	6	A	32	B	S	
64	Is external stream speed important in estimating the performance of ejector jet nozzles?	286 297 145	3	A	37	B	S	
65	The best position at which to put distributed roughness on a model in order to stimulate boundary layer transition to turbulent flow.	(*a*) 122, 128 137, 141 340, 341	4	A	27	A	F	
		(*b*) 112, 137 141, 341	6			A	S	
66	Afterbody pressures on bodies of revolution, with and without wings or fins, in supersonic flight.	(*a*) 122, 240 253, 256 362	10	A	41	A	F	"Afterbody" cannot neatly be expressed in the schedule.
		(*b*) 122 256, 253	11			B	F	"256-interest near the tail" not indexed.
67	The extension of aerofoil theory to wings whose lift is not solely dependent on incidence.	(*a*) 113, 195	112	A	49	A	P	In (*a*) and (*b*) searcher failed to recognise jet flap.
		(*b*) 113, 195 349, 350	40			A	P	
		(*c*) 113, 195 349, 350 145	2			A	S	
68	Criteria determining the length of a finite region of a separated laminar boundary layer in two-dimensional flow.	142 139 137	10	B	32	B	S	
69	The response time of a pressure measuring system embodying long tubes.	417 362 284	6	E	25	B	S	

Question No.	Question	Search plan	Documents retrieved	Indexer	Codes used	Searcher	Success or failure	Remarks
70	The ram efficiency of air intakes let into the sides of a wing or body.	(*a*) 281 282, 285	13	B	42	A	F	"282-Located on aerofoils" not indexed.
		(*b*) 281, 285	40			A	P	
71	Is the temperature recovery factor of a laminar boundary layer altered by separation of the boundary layer?	137 139 142	10	A	27	A	S	
72	Data required from tests employing the NACA Technique for obtaining free-flight stability data from models fitted with an all-moving tailplane which moves automatically between stops under the influence of the aerodynamic forces on it.	(*a*) 112, 115 117, 338 406	0	A	46	A	F	
		(*b*) 115, 117 406	8			A	S	
73	For assisted take-off, what are the advantages of a liquid fuel rocket over a solid fuel rocket?	171 271 377	3	B	32	B	S	
74	What are the handling qualities of the Javelin aircraft at high lift coefficients?	350 378	17	B	38	B	S	The 17 documents will be reduced substantially when the name Javelin (used in indexing) has been represented by a peek-a-boo card.
75	What reports from the Royal Aircraft Factory appeared during the 1914–18 war on aircraft stability theory.	100, 113 114, 117 349, 355	1	A	15	A	S	
76	To what extent may disturbances present in a wind tunnel invali-	(*a*) 125, 137 141, 195	12	B	30	A	F	

	date tests made in the tunnel to locate the position of boundary layer transition on a wing at subsonic speeds?	(*b*) 112, 137 141, 195 392	11			A	S	Indexer omitted to index Mach numbers.
77	What changes in piloting technique are necessary for the safe landing of an aircraft without an undercarriage on an aircraft carrier adapted to receive it?	331 367	4	F	21	A	S	
78	Can Falkner's 9-point method (for calculating the aerodynamic loading on a wing) be simplified for application to a swept wing of low aspect ratio?	113 368 195 206 212	5	E	33	B	S	
79	A method is required for calculating the spanwise distribution of lift on a wing of low aspect ratio at high subsonic speeds, including allowance for the non-linear variation of lift with incidence.	350 362 212 125 123 113	2	A	35	B	S	
80	Wanted: a good theoretical treatment of the downwash field behind a low aspect ratio wing including the behaviour of the tip vortices at supersonic speeds.	113 144 122 143	6	A	24	B	S	
81	Information is required on numerical methods employed in the theoretical determination of the drag of a slender body of arbitrary cross-section.	113 240 352	14	A	16	A	F	"240-body" not indexed. Paper was on area rule, which was indexed.

Question No.	*Question*	*Search plan*	*Documents retrieved*	*Indexer*	*Codes used*	*Searcher*	*Success or failure*	*Remarks*
82	What methods have been tried in flight to improve the damping of lateral oscillation of high-speed aircraft?	112, 193 339, 355 357, 373 405	2	A	33	A	S	
83	Have we any experimental measurements of the influence of aero-elastic distortion on the effectiveness of trailing edge flaps on rectangular wings?	112 374 315 308	3	A	38	B	S	
84	What reports give design charts for the determination of the downwash angle at the tailplane for tapered wing planforms with plain flaps at low speeds.	(*a*) 144, 126 315, 202	2	A	50	B	F	"202-tapered" not indexed
		(*b*) 144, 126 315	11			B	S	
85	Approximate methods to suggest the requirements of an auto-control system for the control of the "long period" or "phugoid" motion of an aircraft.	278 373	9	A	16	A	S	
86	Experimental measurements at supersonic speeds of the flow field in the vicinity of a body of revolution at high angle of incidence.	(*a*) 112, 122 253, 365	0	A	39	B	F	"365-high incidence" not indexed.
		(*b*) 112, 122 253, 128	34			B	S	
87	Calculations to indicate the effect of introducing inertia	(*a*) 109, 302 356, 358	4	C	25	A	F	

	weights and springs into a power-operated control system for the longitudinal control of an aircraft.	(*b*) 278, 302 355, 358	7			A	S	
88	Experimental data on the effects of slipstream on tailplane effectiveness at high subsonic Mach numbers.	(*a*) 112, 144 236, 358 125, 123	5	A	43	B	F	The abstract card, issued with one of the 5 documents, refers to the subject document.
		(*b*) 112, 123 144, 236	10			A	F	"144-slipstreams" and "236-tail-plane" not indexed.
89	Experimental determination of the relationships between turbulent boundary layers on flat plates and cones at zero heat transfer.	112 140 137 333	11	A	35	B	S	
90	Experimental measurements at high subsonic speeds of the pressure distribution over the leading edge of a two-dimensional flat plate with a rounded leading edge.	112 125 333 362	4	A	31	A	S	
91	Wind tunnel tests at supersonic speeds on a canard configuration with ramjets.	112, 122 172, 192 392	1	A	54	A	S	
92	Manometers for automatic measurement of pressures in supersonic wind tunnels.	122, 362 391, 392 417	4	B	23	A	S	
93	The effect of small asymmetries at the nose of a cylindrical body at supersonic speeds.	(*a*) 240, 249 122, 255	3	C	37	B	F	"249-straight parallel sides," and "255-interest at nose" not indexed.
		(*b*) 240 249, 122	43			B	F	
		(*c*) 122 240, 255	10			A	F	

Question No.	*Question*	*Search plan*	*Documents retrieved*	*Indexer*	*Codes used*	*Searcher*	*Success or failure*	*Remarks*
94	Theoretical methods for determining the performance of wing-tip controls.	113, 302 307, 322	3	B	35	A	S	
95	Any information on moving-wing guided missiles.	164, 322 195	5	A	49	B	S	
96	Wind tunnel tests at supersonic speeds on two-dimensional aerofoils.	(*a*) 112 122, 195 338, 392	7	B	34	A	F	"392-wind tunnels" was indexed but not punched.
		(*b*) 112, 122 195, 338	10			A	S	
97	Rapid recording methods for measuring signals from strain gauge wind tunnel balances.	(*a*) 417, 413 391, 416	0	A	18	B	F	"413-balances" not indexed.
		(*b*) 391 416, 417	2			B	S	
98	Design of a six-component strain gauge balance for use in supersonic wind tunnels.	392, 391 413, 122	5	A	23	B	S	
99	Flow at supersonic speeds over bodies of non-circular cross-section.	240, 254 122, 128	11	A	27	B	S	
100	Experimental information about the lift on wings at hypersonic speeds.	112, 120 195, 350	10	A	40	A	S	

TABLE 2 *Classification schedules*

	Code No.	*Times[a] used*
Bibliographic characteristics		
Pre 1935	100	77
" 1945	101	149
" 1950	102	194
" 1955	103	500
Post 1934	104	633
" 1944	105	555
" 1949	106	505
" 1954	107	192
Data/Tables/Methods; for Calculation/Reference	109	58
Bibliography/Survey	110	18
Dictionary/Directory/Nomenclature	111	0
"Experiment"/Reports observation	112	524
"Theory"/Other	113	230
G.B.	114	493
U.S.A.	115	203
Other foreign/International	116	3
RAE/NACA/AGARD	117	621
Firms	118	3
Other	119	77
Mach number/velocity		
≥2.0	120	107
≥1.2	121	176
≥0.9	122	253
≥0.6	123	280
<2.0	124	393
<1.0	125	324
<0.6	126	236
Stationary	127	12
Flow/Fluids studied	128	359
Ideal fluid	129	20
Gas	130	328
Liquid	131	18
Real fluids other than Air/Water (including Rare air/Slip flow, etc.)	132	8
Homogeneous fluid/Total immersion	133	8
Other/Free surface/Mixed flow	134	38
In own right	135	26
In relation to solids	136	279
Boundary Layer	137	94
Flow/Fluids studied (*Continued*)		
Other	138	27
Laminar/Streamline flow	139	51
Turbulent flow	140	59
Transition	141	42
Separation	142	36
Vortices	143	45
Slipstreams/Wakes/Downwash	144	89
Jets	145	27
Blowing	146	16
Suction	147	28
Shock waves/Mach lines/Compressibility	148	76
Circulation	149	4
Condensation/Evaporation	150	4
Convection	151	5
Dissociation	152	1
Gusts	153	16
Heating	154	41
Cooling	155	28
Acoustics/Noise/Sonic bangs	156	6
Aircraft/Aircraft components		
Aircraft "Type"		
Fighter	157	117
Bomber	158	38
Transport	159	22
Personal	160	8
Other	390	38
Shell	161	1
Bomb	162	12
Target	163	3
"Missile"	164	47
"Single" engined	167	90
Multiple engined	168	67
Jet propelled	169	146
Turbojet	170	90
Rocket	171	33
Other/Ram, pulse	172	18
Propeller driven	173	85
Turbo prop	174	21
Other	175	24
Rotary wing aircraft	176	23
Towed	177	7
Sailplane	178	7
Lighter than air	179	3
Kite	180	3

[a] Times used in the indexing of 700 documents.

	Code No.	Times used[a]
Aircraft/Aircraft components (*Continued*)		
Aircraft "type" (*Continued*)		
Parachute	181	6
Other method of propulsion	182	0
Seaplane	183	16
Amphibian	184	1
V.T.O. Type/Flying platform, etc. (not helicopters)	185	4
Composite aircraft	186	0
No wing	187	14
No body	188	5
No tailplane/No foreplane	189	50
No fin	108	20
With multiple fuselage	190	8
With fins other than single tail fin	191	15
With foreplane/Canard	192	6
Components present and/or studied		
Complete aircraft	193	286
Other	194	324
Wing	195	361
Monoplane	196	229
Biplane	197	16
Cruciform, Y, etc.	198	26
Other	199	1
Untapered	201	57
Tapered excluding delta, arrow	202	67
Triangular/Delta/Arrow	203	104
Other (circular, "W," etc.)	204	6
Unswept (inc. 10° sweepback)	205	102
Swept	206	177
Forward	207	2
<40°	208	45
⩾40°	209	108
<50°	210	87
⩾50°	211	43
Aspect ratio <4	212	89
Aspect ratio ⩾3.5	213	67
Aspect ratio ⩾6	214	18
Leading edge sharp	215	15
Wing (*Continued*)		
Leading edge ducted	216	5
Leading edge drooped	217	3
Cambered	218	25
Uncambered	219	11
⩾8% thick	220	56
<8% thick	221	42
Interest in/at/near—wings		
Planform	222	89
Aspect ratio	223	28
Sweep	224	37
Taper	225	19
Other/span	226	25
Section	227	63
Camber	228	25
Thickness	229	34
Other/profile	230	35
Twist	231	22
Dihedral/anhedral	232	4
Leading edge	233	42
Trailing edge	234	61
Tip	235	42
Tailplane/Foreplane	236	79
Fin	237	35
Propeller	238	32
Rotor	239	22
Body	240	139
Pointed nose	241	74
Other nose	243	24
Pointed tail	244	18
Other tail	246	54
Fineness Ratio ⩾6	247	43
Fineness Ratio <6	248	15
Straight parallel sides	249	58
Other	250	42
Rotationally symmetric/Body of revolution	253	80
Other	254	23
Interest in/at/near—Bodies		
Nose	255	17
Tail	256	29
Section	257	10
Sides	258	18

[a] Times used in the indexing of 700 documents.

	Code No.	*Times[a] used*
External store	259	38
Engine nacelle	260	32
Other	261	7
Wing mounted	262	26
Other	263	7
Strut mounted	264	11
Other	265	6
Canopy/Radome/Turret/ "Blister"	266	8
Bomb bay/Release gear	267	6
Guns/Armament	268	19
Escape device	269	2
Aerial/Antennas	270	2
Assisted take-off device	271	7
De-icing gear/Icing	272	4
Fuel/Fuel system	273	3
Fuel tank	274	12
Crew accommodation	275	2
Load/Passenger accommodation	276	3
Dirt excluders/Ventilation	277	5
Automatic control/ Servomechanisms/ Stabilization	278	31
Float/Planing surface	279	12
Undercarriage—Land	280	9
Internal flow	281	96
Located on aerofoils	282	26
Located on bodies, nacelles	283	34
Other/pipes	284	30
Intakes (divide as exits)	285	51
Exits	286	20
Nose/Tail	287	23
Side	288	11
With centre-body	291	21
No centre-body	292	9
Ducts	293	47
Straight sections	294	3
Bends	295	2
Contractions/ Expansions	296	20
Nozzles	297	21
Diffusers	298	18
Other	299	1
Convergence/ Divergences	300	4
Total enclosures	301	9
"Control" device	302	179
Attitude	303	71

	Code No.	*Times[a] used*
"Control" device (*Continued*)		
Drag increase	304	34
Other purpose	305	42
Boundary layer control	306	17
Wing-located	307	104
T.E. region	308	76
L.E. region	309	21
Other	310	11
Fin located	311	26
Foreplane/Tailplane located	312	34
Body located	313	13
Other location	314	5
Flap type	315	93
Plain	316	23
Split	317	18
Other	318	16
Spoilers/Vortex generators	319	15
Fences/End plates/Flow guides	320	17
Auxiliary aerofoils/Slats	321	12
All moving component/ Tip	322	16
Tabs/Balancing devices	323	20
Parachute	324	10
Thrust reversal/Jet deflection	325	10
Other objects		
Engine	326	74
Compressors/Turbines/ Windmills/Fans/Pumps	327	42
Cascades/Stators	328	5
Ground/Runways/Ground Facilities	329	11
Catapults/Launching devices	330	11
Ships/Watercraft	331	6
Man/Medicine	332	6
Flat Plate	333	28
Cylinder—perpendicular to flow	334	3
Particles/Droplets	335	5
Other solid	336	6
Geometry		
3-Dimensional	337	274
Other	338	73
Full scale	339	173

[a] Times used in the indexing of 700 documents.

	Code No.	*Times[a] used*
Geometry (*Continued*)		
Other	340	261
Rough surface/(Surface condition)	341	17
Steps	342	11
Projections	343	21
Fairings/Junctions	344	16
Distortions/Bumps/Indentations	345	13
Recesses/Gaps	346	9
Perforated/Porous/Slotted surface	347	35
Variable shape	348	12
Aerodynamic interest	349	569
Lift	350	193
Drag	352	232
Thrust	353	43
Moments/Stability/Control		
Moments	354	184
Stability	355	176
Control	356	102
Lateral	357	108
Longitudinal	358	209
Directional	359	88
Hinge	360	30
Damping	361	44
Pressure/Velocity distribution	362	192
Loads/Loading	363	83
Interference/Relative position	364	89
Stalling/High incidence	365	53
Sideforce/Sideslipping	366	34
Spinning	251	27
Landing/Take-off	367	56
Level flight	407	28
Diving/Descent/Gliding	368	34
Climbing	369	24
Turning	370	21
Hovering/Flapping	242	14
Separating/Jettisoning	371	13
Buffeting	372	15
Flutter/Oscillation	373	48
Aeroelasticity	374	23
Magnus effect	375	0
"Area rule"	376	10
Operation/Design	377	92
Load/Capacity/Performance/Range/Speed	378	79
Operation/Design (*Continued*)		
Economics	381	16
Reliability/Maintenance/Safety/Accidents	382	25
Civil operation	383	17
Military operation/Strategy/Tactics	384	45
Ballistics interest	385	20
Structures/Materials interest	386	47
Mechanical interest	387	35
Electronics/Radar/Radio	388	24
Mathematics	389	51
Principal test equipment/Technique		
Described	391	133
Wind tunnel	392	261
Continuous/Fan drive	393	45
Intermittent	394	8
Shock tube	395	4
Special purpose tunnel/Test rigs	396	23
Spinning tunnel	397	10
Water tunnel	398	4
Ballistic range	399	4
Analogy/Simulator	400	4
Ditching tank	401	2
Towing tank/Moving channel	402	12
Models mounted on aircraft	403	1
Tracks/Sleds	404	1
Free flight—Piloted aircraft	405	92
Free flight—Models/unpiloted	406	41
Components	408	24
Power/Flow production	409	15
Other	410	13
Auxiliary Apparatus/Technique	411	58
Model Design/Construction	412	49
Balances	413	18

[a] Times used in the indexing of 700 documents.

	Code No.	Times[a] used
Principal test Equipment/ Technique (*Continued*)		
Scale Effect/ Corrections/ Calibration	414	31
Telemetry	415	8
Data Handling/ Methods/Computers	416	10
Flow visualization	252	51
Instruments	417	59
Combustion	418	17
Meteorology	419	8
Optics	290	4
Photography	165	11
Total postings in indexing 700 documents		20,270

[a] Times used in the indexing of 700 documents.

SUMMARY OF DISCUSSION

Professor Eric de Grolier, Panel Chairman for Area 4, began the discussion of comparative characteristics of existing systems by drawing an analogy between the evolution of systems for transportation over water and those for scientific communication. Man began, in the one case, by such "pedestrian" methods as wading and swimming, followed by the use of artifacts, the development of galleys and sails, and eventually the steam, diesel, and atom-powered ships of our latter day. Similarly, systems for communication evolved from oral methods through the use of artifacts such as books carved in stone, through the development of printing to mechanical methods of handling information and again, with recent rapid progress in new technologies, to the use of electronic computers for information selection and retrieval. If such use of computers is rightly viewed as analogous to the exploits of the Nautilus, immediately the question of the basic objective is brought to the fore: Should we use the Nautilus for crossing a ford that we could wade, on the one hand, or should we attempt to *swim* the Arctic Ocean, on the other?

Other useful bases of comparison in the analogy include the fact that as the artifacts used become more and more complex, but also more and more efficient, their investment costs become higher and higher, yet their operating costs become smaller and smaller for a given workload. There is also the important point of similarity that the new machines and methods have never entirely suppressed the old. They have supplemented them, replaced them for certain tasks, and, significantly, have permitted other and entirely new tasks to be undertaken.

In intellectual as well as in physical systems of communication there are different fields of application of each means or tool that is developed and there are different thresholds which separate these fields as a function of the difficulty, complexity, and size of the task to be accomplished.

As the various systems for scientific communications have evolved, there has been a continuing coexistence of the products of the different stages of evolution: correspondence between working scientists, publications in primary journals, the emergent bibliographical aids, and, presently, the multitude of large and small documentation centers with at least the three functions of accessioning, processing, and redistributing material.

Since 1939, there have been various attempts to centralize or unify documentary efforts on a national basis, ranging from such efforts as the French National Center for Scientific Research to the All-Union Institute of the USSR. Worldwide, there is somewhat analogous evidence of the will to cooperate, to coordinate, yet this may be taken in the sense of a French saying to the effect that when one does not really know how to organize, then one coordinates.

The point is, however, that what one wishes to coordinate, centralize, and organize, embraces the gamut of information selection and retrieval systems at all levels from the small but meaningful index maintained by the individual bench worker to that represented by huge national centers such as that of the USSR. What is good as documentation practice for the individual scientist is not necessarily good, for instance, for the French Atomic Energy center where documents accumulate at the rate of more than 100,000 new accessions per year. Evaluations of different proposed systems cannot reasonably be made without differentiation between such varying requirements.

A second necessary objective is to recognize the different fields of scientific information: for example, are the social sciences, even the science of religion, to be considered as within this scope? If so, the user requirements of the social scientist may differ from those of the physical scientist, as also those of the chemist differ from those of the physicist or the engineer. Must we not recognize that each field has its own particular needs and particular requirements? Finally, there are the different types of information for different communication purposes, from horizontal communication between scientists on the same level, to vertical communication from the pure scientists to the man on the street, passing through the applied scientist, the technologist, the shop foreman, and at the other extreme, to researchers in quite different disciplines.

Such compelling, and challenging, reflections set the stage for the ensuing discussions, both general and specific. For the panel, Mr. Eugene Wall first discussed some general terms on which comparisons of existing systems might be based. Two sets of variables must either be controlled, or their effects on the system being evaluated or compared must be quantitatively known or measured. These are: *1,* the external or environmental variables, and, *2,* those peculiar to the internal system. It is the necessary interaction between the two types of variables which must determine both the technical and the economic effectiveness of any system. Thus it is doubtful that any one panacea for all information handling problems is realizable—that any one combination of internal system characteristics can be optimal for varying combinations of environmental factors. We should seek, therefore, for any underlying set of principles by which we can adjust internal or system variables for optimal interaction with environmental requirements. Perhaps this can best be achieved by the improvements of communication between the user and the system, and of the match between the typical question and its answer, and not by seeking to improve the match beyond an economic point.

In seeking such an improvement, various technical problems must be solved. First is that of viewpoint, since the experience of various individuals is usually such as to exclude universal agreement on the assignment of any given concept to any one logical class. The second problem is that of generics or breadth or narrowness of viewpoint. The sought-for answer should be available at any equal or lower generic level of any of the family trees that stem from a given concept. The third problem is that of semantics—the meaning of words or relations between concepts and their symbols. Different words often mean the same thing, or identical words different things. A fourth problem is syntactic, that of the ordering of words or concepts in such a way as to express the direction of relationships between them.

Mr. Wall suggested the need for a *continuous* rather than a discontinuous solution to these four problems, that is, a solution which can by increasing costs incrementally thereby also incrementally increase the effectiveness of the match between question and answer. Only by the development of principles which enable the solution of these problems in a continuous fashion can a basis be found for evaluation of interactions of information systems with their specific environments. General principles yielding continuous solutions to such linguistic and semantic problems would provide a smooth curve along which to plot cost versus effectiveness. Small steps along the curve would then focus optimum points for particular systems and their particular environments.

All facets of documentation from the origination of the information to its ultimate use must be considered because they interact with each other as well as with the environment. The cost-benefit relationship must be determined, not only for a single task, such as indexing, but for the system as a whole.

Mr. Wall then outlined the principal environmental factors to be considered. First is the value of the information in the collection. The more valuable the information the more often it will be referred to and, conversely, the lower will be the accession rate since low value material probably will not be retained. This results in a relatively high ratio of references to accessions, where it may be profitable to expend effort to put things away well so that the subsequent work of reference is speeded. Contrariwise, with a low output-to-input ratio it may be best to search harder rather than to index better.

A second external factor is the relative transience of the material to be filed. A third is the number of items in the file, requiring, with increasing volume, an increasing specificity of questions asked and a correspondingly greater effort on input in order to ease the input task. A fourth consideration relates to the number of potential customers, and the fifth to the types of question asked.

To summarize, there is no one solution adaptable for all environments—no one right tool for all tasks. Secondly, various possible systems must solve the technical problems of viewpoint, generics, semantics, syntactics. A set of general principles for the solution of these problems must surely exist. Mr. Wall further suggested that comparisons among systems will be valid only as the solutions to such problems are continuous and as environmental factors can be generalized in the comparative evaluation. Fourthly, he declared that the logic of the system should dictate machine design. Finally, he stressed the opinion that all facets of a documentation system must be considered as well as a number of important environmental factors.

Dr. Brian C.Vickery was the next panel speaker. His comments reinforced the view of the necessity for consideration of systems as wholes. In designing a system for a particular situation, we must take into account the total environment, the needs of the user, the costs and benefits of each step in the whole documentary operation. The environment determines what a retrieval system is required to do—its basic function. If a new system is to be designed to carry out certain specified functions, then we must know how those functions and the structure of the proposed system are related. Comparisons of one system *vis-a-vis* another may show one better suited to a particular environment than the other, but we should know more than that, we should know *why,* for what purpose, one system does or does not function better than another. If there are different ways of carrying out a certain function then these ways should be intercompared to decide which does a given job most efficiently. It may be necessary to construct a series of model systems in which only one feature is varied at a time.

Of the various internal system variables suggested in the introduction to Area 4, only a very few have been tested under any circumstances for any systems at all. Tests reported in the papers for Area 4 provide, as yet, no basis for generalization. Indeed, the tests of Mr. Cleverdon, although designed to study variable features of the over-all problem, underline the truth that no single variable factor can be considered in isolation.

But the whole point of intercomparing the characteristics of existing systems should not be to show that one system is better than another. On the contrary, a principal concern of conferences such as this, and of all who are interested in the field of documentation, should be to find out how we can design new and improved systems that will aid all in the most efficient way. We want to be able to generalize from comparisons, and perhaps to find systems, as a result of comparisons, which will be better than any system we now have.

The comments of the next panelist, Dr. C.J.de Haan, re-introduced the perspective of looking at the common factors in conventional systems and in modern systems using machines. In both, the objective is to reduce work, namely, retrieval work. The needs of the user, in particular differing needs, such as those of the research scientist in contrast to those of the applied scientist, must be taken into account. Economic factors are basic, since we are never justified in working out new systems whose costs in labor, investment, and effort, exceed what could be done with similar results by a more conventional system. What the systems should be depends largely on the information that the users of the system are likely to require.

In order to meet the objectives for which machine systems are being developed, three operations are of great importance. First, the analysis of each document in the file to be searched in terms of all features relevant for future retrieval. Second, the drawing up of dictionaries and standardized terminology for such retrieval features. Third, the translation of identified retrieval features into a language suitable for machine use.

First, then, the analyses of documents must be carried out in the most accurate way possible. Each feature overlooked or improperly measured must lead to incomplete or erroneous answers when machine searches are made. This task of analysis is both the most important and the most expensive part of mechanized search systems. It requires a great amount of labor on the part of technically qualified persons. Thus, in machine searching, the exercise of human intellect and judgment cannot be eliminated. It can be concentrated. It can be condensed in such a way that the work need not be repeated for each question to be answered, but at least once the intellectual effort must be expended. A pertinent criterion for determining applicability of machine systems is therefore the question of manpower investment necessary to select and define essential features—those that will be asked for by users of the system—and to trace these features in the analysis of the documents.

The remarks of Dr. John W.Kuipers again stressed the difficulties of carrying out the mandate of the Area 4 panel, of separating the considerations proper to this Area from those of Areas 5 and 6, and of attempting to fix criteria for the evaluation of systems. He suggested that the basis for these and similar difficulties, not only in Area 4 but throughout the Conference discussions, may lie in a tendency to concern ourselves with secondary questions rather than with primary ones. Such matters as specific machines and recording media, economic questions, and notation schemes are secondary to basic questions concerned with the over-all problem of communication. What is the over-all task to which we set ourselves? The system, the integrated set of procedures by which specific tasks may be accomplished, will necessarily be complex. The system should take into account all required functions, including those of generating messages, recording, processing, disseminating, transmitting, storing, searching, collecting, and using. The system must take into account the dynamic nature of the situation in which it operates. It should take into account that there are many individuals and many organized groups attempting to communicate. Certain parts of the ultimate over-all system are vague at this time, but there is much that can be stated and clarified. We can hope to fill out the picture only as we begin to get better answers to other primary questions: What is the nature of the messages to be carried in the system? How do we get from expressions in natural language as used by human individuals to whatever information units will flow in the machine part of the system? If there are limitations to be imposed on the messages the system handles, what are these limitations and why do we accept them? If we can come to some general clarification and agreement on questions such as these, we may hope to proceed more profitably with matters of secondary concern such as are involved in comparative characteristics of classifications and of machines.

This part of the panel discussions concerned with the general philosophy of evaluation was closed with brief remarks by Dr. Jacques Samain urging greater standardization in the preparation and presentation of documents, and by concise, practical observations by Dr. Hans Selye based upon experience with a system that has been used for over a century.

Dr. Selye suggested first that in the development of any system a primary objective should be the molding of that system by the practice gained in its use, rather than by theoretical presuppositions or by undue concern with how many hits or errors the system makes. It is a matter of giving higher credit in any system to the finding of such information as is particularly difficult to find. An occasional positive hit, the tracing of a very important but rare document, might be very much more important to the scientist than a high average of successful findings of information on a statistical basis.

Dr. Selye also re-emphasized the basic importance of the intellectual task of taking a document and extracting from it that information which will be really useful to later users. He therefore urged the desirability of establishing better cooperation between those who know the mechanical and linguistic aspects of organizing information for search and retrieval and those who actually use such information. Suggestions for the organization of working parties for detailed consideration of practical problems and tasks were endorsed by the chairman, who then opened the discussion to the floor.

Discussions from the floor following the general comments by panel members included clarification of certain specific points raised by the papers or made by the panel members. Mr. R.A.Fairthorne offered comments on the apparent misunderstanding as to the extent to which machines could perform intellectual work, stating that what the machine is meant to do, just as a book is meant to do, is rather to prevent the repetition of the same intellectual work over and over again.

Dr. M.J.Taube suggested that the question of the right tool for the right job, a best system for a particular situation, depends upon the assertion of a theory of search, storage, and retrieval.

Questions were raised with respect to suggested criteria, such as the relationship of input-output ratio to the cost-benefit factor; to possible reasons for differences between the Herner and Herner results, using user language, and those reported by Dr. Whaley; to possibilities for converting any one system to any other; to problems in planning and carrying out of tests and experiments, such as those of Mr. Cleverdon, and to re-emphasis of the fact that a feature of any system is that we cannot get out what has not been put in. Mr. Farradane offered an emphatic plea for greater application of scientific method in testing and experimentation, and for realistic objectives in terms of system efficiency. It is time, he suggested, that we adopt a truly scientific approach to what must become, if it is not yet, a scientific subject area. We need proper experimentation, proper ideas of isolating different parts of a problem, and of testing them properly in accordance with well-defined, objective procedures. Dr. E.F.Moore suggested that an important factor in the success of systems in use today may well be the interest and abilities of those who do the indexing and abstracting and of those who are in charge of the operations, rather than the performance characteristics of the machine or system itself.

These comments from the floor provided ready transition to the next topics discussed by the Area 4 panel—the problems of attempting actual intercomparisons and the selection of specific criteria for evaluation. Mr. Gerald Jahoda reported on the results of a questionnaire survey of some 39 correlative or coordinate indexing systems now in operation, mostly located in the United States. These systems are relatively small. Over 80 per cent of the installations participating in the survey reported collections of 20,000 or fewer documents. The systems are not extensively used, since 60 per cent reported 200 or less inquiries per year. Finally, there was a strong indication from the questionnaire responses that in the large majority of the installations more traditional methods such as alphabetical subject indexing or conventional classification systems might have been used to equal advantage. If so, the question is properly to be raised as to whether we should not be less extreme in formulating information retrieval requirements. At the present time, neither correlative indexing systems nor conventional indexing techniques prove particularly adequate for doing an increasingly formidable job. Should we not then ask ourselves if there may be quite different answers possible to problems of information storage and retrieval and proceed to research on such other solutions? On this last point, another panelist confirmed the need to explore alternate possible solutions but interposed the suggestion that since the less conventional systems are very new they should not be judged prematurely, as though, in an example accredited to Sir Winston Churchill, we were to ask, “What good is a new-born baby?”

In the following remarks, Mrs. Claire K.Schultz discussed the practical problems in isolating, defining, and applying comparative factors. An experiment in intercomparing two closely similar mechanized systems, dealing with the same general subject matter, led to the conclusion that the systems should be studied as whole systems but also to the difficulty that all the necessary interrelationships between the functional parts of a system are not easily traced. One of the important lessons learned from the study was the need to use real questions, or real questioners, in evaluating output.

Some of the factors considered included the following questions: How quickly must the answer be found? For what purpose is the answer wanted? Is the client a student, a specialist in the field, an administrator, someone who needs to document a legal dispute? Is the answer to be limited, for example, to review articles or to a particular language? What form of output—bibliography, journal, microfilm, digest of information contained in the selected references—is desired? What is the usual depth of indexing, bearing in mind that the number of descriptors used may not be directly correlated with this since various depths may be represented in several fields that may be covered in any one document? What proportion of the indexable material in the fields covered is indexed? What types of errors can be made in input and what are the costs of errors that are undetected?

These and other critical factors were discussed with emphasis upon time and costs of necessary staff effort. For example, if the user would allow two weeks to elapse before obtaining an answer to his question this would not necessarily mean that the staff should use two weeks to find that answer. Again, the time spent screening to reject or eliminate material should be accounted for as well as the time spent on actual input of selected material. How much professional time, and at what cost, is involved in entering selected documents into the system? How costly in manpower, money, and materials are checks that are made to prevent or reduce the introduction of errors?

After these specific examples of critical factors, Mr. C.D.Gull commented that follow-up of criteria suggested in the papers and in the discussions would be desirable both in terms of expansion and refinement and in terms of the design of suitable tests and experiments to apply them. He then took the

Wright-Wilson paper as a point of departure for the elaboration of further critical considerations. Can we, for example, regard a system as satisfactory if the user must discard a high proportion of the output product as irrelevant to his query? If a relatively large number of descriptors or index headings are assigned per document, why is it that significant points are still missed? To this latter question the senior author rejoined, thereby reinforcing a major theme of the Area 4 discussions, that the failures were in the intellectual analysis, not in the mechanics of the system used.

The final panel member presentation was by Dr. Robert M.Hayes, who first re-stated such key questions as that of whether systems were to be considered as equipment, as coding or classificatory schemes, or as total organizations of resources and procedures to meet specified objectives, and the question of overlap of Area 4 with other Areas. He pointed out that both the development of a new system and the comparison of two existing systems involve the trade-off between a number of factors and that criteria typically have practical meaning only when they are considered within the framework of a general theory. He raised the point of necessary objectivity in the development and application of criteria for comparison and questioned the extent to which tests such as those of Miller, Cleverdon, and others produce results that can be extrapolated to a general situation. Within the framework of a good general theory, however, tests can be used to establish basic parameters for given systems within a controlled environment.

Continuing the discussion of specific possible criteria, Dr. Hayes suggested first the desirability of interdisciplinary meetings where users, librarians, systems analysts, and engineers might outline the relevant factors of particular concern to each. He then outlined a number of possible qualitative factors for consideration, including those that reflect usage requirements, such as the purpose of the system and the level of complexity of the material to be filed and searched; those that are involved in organization, data format, code notations, arrangements of the file; those that involve the organization and performance of equipment; and finally those that involve elements of design.

The floor was then opened for general discussion, with comments on the use of large versus small computers, the suggestion that a most significant criterion in the evaluation of systems is that of the capacity of the system to change, and discussion of the possibilities for abstracting selected operations from a total system in order to measure them objectively.

In closing, the panel chairman drew three principal conclusions from the discussions. First was the implication that it is necessary to plan and conduct many more experiments than have been tried hitherto, in order to test conditions under which comparisons may be meaningful and to define and improve the criteria suggested for application. The second conclusion drawn was that the survey that had been originally proposed but not carried out prior to the Conference, could and possibly should be conducted, drawing upon the actual experience of this Conference and following methods such as those of descriptive natural history. The third conclusion was that special studies of subjects not specifically included in the papers or discussions, especially matters of intellectual categorization or codification, should be given attention.

MARY ELIZABETH STEVENS, *Rapporteur and Area Program Chairman*

ERIC DE GROLIER, *Discussion Panel Chairman*

AREA 5

ORGANIZATION OF INFORMATION FOR STORAGE AND RETROSPECTIVE SEARCH

Intellectual problems
and equipment considerations
in the design of new systems

AREA ORGANIZATION

Authors of Papers

Members of Discussion Panel

Chairman: GILBERT W.KING, I.B.M. Research Center, Yorktown Heights, N.Y.

C.W.CLEVERDON, The College of Aeronautics Library, Cranfield, Bletchley, Bucks, England

G.MALCOLM Dyson, Consultant in Chemical Engineering and Scientific Documentation, 1 Burton Walks, Loughborough, Leicestershire, England

J.E.L.FARRADANE, Scientific Information Office, Tate and Lyle Research Laboratories, Keston, Kent, England

E.F.MOORE, Bell Telephone Laboratories, Murray Hill, N.J.

JOHN J.O'CONNOR, Remington Rand Univac Division, Sperry Rand Corporation, Philadelphia, Pa.

J.W.PERRY, Center for Documentation and Communication Research, School of Library Science, Western Reserve University, Cleveland, Ohio

S.R.RANGANATHAN, University of Madras, Bangalore, India

J.ALBERT SANFORD, Office of Central Reference, National Security Agency, Fort George G.Meade, Md.

DON R.SWANSON, Control Systems Division, The Ramo-Wooldridge Corporation, Los Angeles, Calif.

PROPOSED SCOPE OF AREA 5

SYSTEMS AND PROCEDURES for storing and searching recorded knowledge are designed for the purpose of making the information available when needed. Such systems are often rendered complex by the requirement that information is to be drawn from a variety of sources and also by the fact that in some instances systems must provide both for the recall of individual facts and for their effective correlation. This area is concerned with the design of effective systems, with the problems encountered in processing recorded information for subsequent search, and with the possibilities of using machines or devices in the processing, storage, and search operations.[1]

The effective organization of recorded knowledge requires more understanding than we now have of the relationships between scientific concepts and the words and other symbols in which they are described or expressed. The processing of information for retrospective searching must also take into account various practical considerations such as the purpose to be served, the capabilities of equipment that is already available or that could be constructed by exploiting present-day technological possibilities, and the costs involved in applying the various methods of processing information.

We must consider not only the effectiveness of traditional procedures, such as alphabetical indexing and hierarchical classification, but also recent experimental work and research trends directed toward the development of new or modified methods of indexing or organizing material as well as the development of equipment to aid in the rapid processing and searching of information.

PRINCIPAL SUBJECTS FOR DISCUSSION

The principal subjects for discussion might be grouped under the following headings:

1. Linguistics—Structure and Meaning;
2. Problems Encountered in the Design and Application of Systems;
3. Application of Machines in Storage and Search Processes;
4. Comparative Evaluation of Experimental Systems.

[1] The proposed scope of the Conference Area, as shown here, was prepared during the Spring and Summer of 1956 and provided to all potential contributors as a guide to the aims of the Conference.

1 LINGUISTICS—STRUCTURE AND MEANING

The relationships between scientific concepts and the words and symbols in which they are expressed may be highly complex because symbols are entities in their own right and frequently more than one concept is associated with a single symbol. It is necessary to consider the problems involved in defining the meanings of words and also the relationships between words whose scope of meaning is narrow and words of more generic significance. The possibility of making use of generic concepts in correlating recorded information is particularly worthy of attention. The knowledge and methodology of linguistic scientists might be examined also for contributions in the effort to deal with the problems that occur in attempting to organize information. It may be helpful to study the devices used in natural language for indicating relationships between words—devices such as word order, directional prepositions, and various types of syntactic structure.

2 PROBLEMS ENCOUNTERED IN THE DESIGN AND APPLICATION OF SYSTEMS

The question as to whether subject matter is to be recorded for storage in the form of a total text, an abstract of that text, the key words selected from the text or some other form depends upon the type of service to be provided. Whether the information to be stored and later searched involves ideas or notions on the one hand or established facts on the other will influence the design of methods for organizing the subject matter and the selection of equipment to be used with the storage and search system. The problems encountered in designing and using the following types of indexing and classification schemes and the capabilities and limitations of the various schemes should be explored.

A Conventional Indexing

The word “index” according to Webster's dictionary, is sometimes used to denote “a directing sign or instrument, that which points out; that which shows, indicates, manifests or discloses; a token or indication.” Taken in this broad sense, the establishment and use of an index is essential in the retrieval and correlation of recorded knowledge.

In a narrower sense, the word “index” may be used to denote “a list, usually alphabetical, of topics, etc., in a book, giving the numbers of the pages where each subject is treated, and commonly placed at the end of the volume;” also, “a similar list of references to a series of volumes, or literature in general, usually printed separately.”

The basic question of how words are most effectively used to provide "directing signs" and also the possibilities of using equipment to facilitate the preparation and use of indexes should be considered.

B Coordinate Indexing

When a system of coordinate indexing is used, the key words are used as "coordinates" and in the searching process it is possible to search for any desired combination of these "coordinates" or index entries. Scientific inquiries are usually expressed in terms of multiple concepts arranged according to the inquirer's particular needs of the moment.

Various devices for storage and search can be used effectively with methods of coordinate indexing. The work of combining categories as needed to satisfy an inquiry is left to the machine or device. The high speed at which desired patterns can be detected is the basis for the usefulness of such equipment.

C Indexes that Specify Relationships

Most conventional and coordinate indexing systems are based on key words gleaned from the scientific records to be stored. A combination of the key words and the explicitly-stated relationships between the words forms a telegraphic style abstract that gives the subject content of a document at a glance. The merits of such abstracts for indexing purposes, the methods of preparing and coding the abstracts, and characteristics of equipment needed to conduct searches when information is so organized need to be explored.

D Classification

The broad purpose of classification is to draw together those records that have common features of subject content. The growing number of scientific records has required the use of increasingly narrow subject divisions in classification schemes, meaning that finer distinctions must be made in analyzing subject matter and that the task of identifying documents with common characteristics has become more and more complicated. Broad classification schemes that attempt to cover all knowledge are being supplemented or even supplanted by specialized classification systems covering relatively narrow areas of subject matter. The use of carefully prepared specialized classification schemes or of a multiplicity of such specialized schemes may be very effective in organizing information for subsequent searching.

3 APPLICATION OF MACHINES IN STORAGE AND SEARCH PROCESSES

The type of information stored and the types of questions to be directed to the stored record determine the methods to be used and the kind of equipment that will be selected for a given system. It is sometimes desirable to perform some operations on the information stored in the record, such as determining connections or relations between items. On the other hand, it is sometimes sufficient merely to recognize the presence or absence of items of information or of combinations of items recorded as such. Consideration needs to be given to general characteristics, capabilities, and limitations of the various types of equipment, including examples of such equipment now available which have not been exploited to full advantage.

Reviewing the stored record, rearranging the subject matter, and continually expediting searches can be accomplished mechanically as machines "learn" by experience. Feedback can be employed to establish search priorities, optimum search patterns, and so forth. The mechanics of conducting such maintenance steps should be considered here, and the experience of others who use automatic data-processing equipment should be examined for applicable techniques.

During the searching operation, it is possible to use the answers that first result from the search to direct further searching. This can be a simple process of counting the items yielded by various modifications of the original question, and analysis of the results of such counting to guide additional searchings. On the other hand, auxiliary memory devices can be used with the searching equipment—devices capable of looking for associated items, of recognizing the frequency of occurrence of such associations, and of changing the direction of the search while it is going on. The techniques of searching, the development of search strategy, and the practical problems of conducting searches should be examined here.

In all the discussions of this problem area, emphasis needs to be placed on total information systems, methods as well as equipment. Experiences within particular situations need to be discussed in terms of their applicability to the problems of other situations.

It should be borne in mind also that machines can be used most effectively to prepare statistical analyses of a given body of information and that such analyses could be most helpful in determining the best method for the mechanized handling of that information. First a statistically meaningful sample must be selected from the total body of information. It should then be analyzed in various ways for its characteristics, such as the frequency of occurrence of certain kinds of information or the frequency of occurrence of certain combinations of items. The methods of employing present equipment, or the requirements for new equipment, to conduct such statistical analyses have not yet been fully recognized. Research in the field of mechanical translation offers one example of the use of such analytical methods in designing mechanized information systems.

4 EVALUATION OF EXPERIMENTAL SYSTEMS

It appears that persons and groups who are engaged in the development of systems embodying new principles for organizing subject matter could effectively compare their systems by applying them to a common set of documents. This may be a small sample, to be agreed upon by the cooperating investigators. It is suggested that the papers from the August 1955 Conference at Geneva, on peaceful uses of atomic energy, might constitute a useful sample. Use of the same sample for all systems would make comparison and evaluation easier. The systems should be discussed in terms of:

1. Size of the file to be covered;
2. Rate of growth of file and system;
3. Range of inquiries to be serviced, or the purposes to be served;
4. Range of subject matter to be covered;
5. Kinds of concepts to be represented;
6. Specificity and type of analysis;
7. Personnel required to do the analysis;
8. Cost of processing information and conducting searches;
9. Reliability of results, or probability of retrieval;
10. Form of system.

It is recognized that discussions in this problem area cannot be as factual as the descriptions of systems in actual operation could be. But the discussions should be practical and point out capabilities and limitations of the systems insofar as they can be determined.[2]

5. Papers on the bearing of studies in linguistics on storage and search systems, papers on problems encountered in the design and application of systems, papers on the application of machines in storage and search processes, and papers evaluating experimental systems are especially invited.

[2] No such study was undertaken, as far as is known.

The Basic Types of Information Tasks and Some Methods of Their Solution

V.P.CHERENIN

GENERAL INFORMATION ABOUT RETRIEVAL TASKS

THE POSING OF THE TASK

The defined set of information among which it is necessary to carry out a search is represented in the form of a universe *E* of individual elements of information e_k (1–3). The answers to questions asked are particular subsets of *E*. The volume of *E* can grow by means of the addition of new e_k, that is, *E* is considered as defined only for a definite interval of time. Data from the elements of information are linked more or less closely to one another, since the condition is usually assumed that they deal with one and the same "subject" (concept, experiment, process, substance, animal, person, etc.). On the basis of considerations which will be set forth below, the universe *E* is, as a rule, made up of homogeneous e_k. Proceeding from the defined set of information, the scope of the questions asked is also determined more or less accurately. In general, this does not mean that all the possible questions are previously known, but that for each question it is usually known whether or not it can be asked. The general conditions for the selection of e_k for the questions asked are also known.

The solution of the task "Does it answer the question asked?" for each e_k, generally speaking, does not evolve clearly from the information contained in the e_k and the question. It can require either a processing of that information according to definite rules or the referring only to information from other e_k (if the set of the information is closed), or even the referring to a more general set of information than *E,* or, finally, it may go beyond the limits of searching for certain information and may require the carrying out of scientific research work. Thus the work of carrying out searches is in its immediate, original form one of the varieties of mental labor requiring not only great erudition and memory in order to cut down the number of references to the data which

V.P.CHERENIN All-Union Institute of Scientific and Technical Information, Academy of Sciences of the USSR, Moscow.

are not included in the information element being immediately considered, but also great art in processing the data from the e_k being considered and from the question according to definite rules. In order not to resort repeatedly to such a complex and slow process during each search (it is assumed that searches will be made many times), it is obviously desirable either to record, by some method, the results of searches previously made, or to create a special information language and to use it to record the content of the information in all the e_k, as well as the content of the questions asked, in such a form that will make it possible, to a greater or lesser extent, to automatize the process of solution, without at the same time leading to an inadmissibly large increase in the volume of initial data or time expended for the solutions.

The first method of solution cannot be applied to all possible questions (of the defined scope), if there are very many of them, and therefore the entering of all old and all newly incoming e_k is done only for a limited number of rather general questions or for more limited, more important questions. This work can be done periodically (as a sufficient amount of new e_k is accumulated) in the form of bibliographic reviews, reference books, tables, etc., prepared by specialists on the individual questions. This work is done continuously as new questions come into the information section, by having the section personnel examine all the existing material, with attempts being made here to use in part the results of searches made by the users themselves, either on the basis of the content of the requests (library) for sets of e_k which they found as answers to the questions in which they were interested, or on the basis of lists of sources which they cited *(e_k)* in articles which they published (citation indexes (4)). Subject classifiers also carry on continuous work in recording each incoming e_k on a more or less established list of questions according to subject headings, from which subject indexes are then prepared. A secondary information task, however, develops as a result of this work: since the answers are found not for all, but usually only for a small number of all possible questions, it is often necessary to find for an arbitrary question (once again, from the definite scope) all the similar questions "corresponding" to it from among the already processed number of questions, which thus form a new set of elements of information E. For example, a system of cross references in the subject index is planned for this purpose. The solution of this secondary task has, as a rule, already had to be done (with the large number of all possible questions) by the second method, which is the only one that we shall consider in the future.

INDEXING

With the second method of solution the content of the information in each e_k and the content of any question asked is, in general, approximated more or less on a one-to-one basis by the set of standard semantic units—characteristics —with an indication of the standard synthetic relations between the latter (1–3). Such a process is usually called indexing. By synthetic relations we understand the connections which can be established between the characteristics directly on the basis of the content of the e_k being indexed or the question. A particular degree of accuracy and one-to-one correspondence (the most that can be permitted is a few approximations) of indexing depends upon our knowledge of the e_k and the question (it is impossible to approximate something with greater accuracy than the accuracy of the assignment of that something), upon our possibilities with respect to the scope of the work necessary to carry out a particular approximation and the time, complexity of equipment, and volume of entries necessary for solution by means of comparison of approximations of the particular accuracy. It is also necessary for the degrees of accuracy of approximations of e_k and the question to be approximately identical. Therefore, it is often necessary, even for tasks in which a high degree of approximation is applicable, to limit oneself to lower degrees of the latter.

The demand for a one-to-one correspondence of approximation is, I dare say, more important than the accuracy, if by a shortcoming of accuracy we understand the searching for superfluous e_k which do not answer the question asked, but not the loss of necessary e_k, since when the one-to-one correspondence of approximation is violated, necessary e_k will deliberately be lost. Therefore, it is sometimes better to approximate e_k and the question less accurately, but on a more nearly one-to-one basis.

In general it can be said that the content of each e_k and question must be approximated on a one-to-one (or almost one-to-one) basis by the logical function of logical variables, after which the selection conditions to be satisfied must be checked by means of definite logical operations on the functions of e_k and the question. In this process the conditions for selection can in general be approximately described as the requirement that the function of the question be the logical consequence of the function of the e_k. We shall meet an illustration of this approach below.

Thus, in the final analysis the special information language must be a variety of that “language of science” about the creation of which many scientists have dreamed. Without posing the task too broadly, but remaining on the ground of reality, we shall limit ourselves in the future only to the simplest degrees of approximation which are applied in practice and which require a not-too-complicated information language, which is so called, rather than the “language of science,” in order to emphasize its simplicity and its immediate practical applicability. It is on the basis of these degrees of approximation that the basic types of information tasks will be introduced.

Getting somewhat ahead of ourselves, it can be said that with an increase in the number of the type there is an increase in the complexity of approximation and the entire information task. Therefore, it is obvious that for any information task it is necessary to select the lowest type providing for a one-to-one correspondence and the minimally acceptable accuracy of approximation. The acceptable accuracy is determined on the basis of the practical requirements, and the possibility of carrying out the particular one-to-one correspondence and accuracy of approximation, on the basis of the content of the e_k and the questions and on the basis of the results of experimental searching. The greater the accuracy of approximation, the more difficult it is to observe its one-to-one correspondence. This fact, which is well known to any classifier, also forces one to get by with the minimally acceptable accuracy.

The simplest method of indexing is the direct processing of the content of the e_k and the questions according to definite rules without referring to any other data (with the exception of those rules). With the least accuracy of approximation this comes down to extracting from the content itself for each e_k some one particular characteristic (or several independent characteristics) used as the standard name for that e_k (type I). Examples of such names could be: the last name of an author or the name of an article, the name of a substance in some system of designations, or the principal subject of an article recorded in a definite form. When this degree of accuracy proves to be inadequate or the names obtained cannot be considered as standard ones (and it is difficult to standardize them), then the content of the e_k and the questions is semantically subdivided into smaller units, for each of which the name described above may prove possible, on the basis of the characteristics which are encountered in them and which are accepted as the standard. This possibility is facilitated by the lesser concreteness of each part and, as a rule, the lesser number of all possible parts as compared with the total number of all e_k or questions, and this makes it possible to standardize these parts more easily. This subdivision must be done on a one-to-one basis and must be carried out either according to definite sections (type II), or with the aid of other sufficiently unambiguous rules expressed with the aid of a standard structural formula for content. Since it is assumed that e_k is taken as the sets of information which are not able to be subdivided into completely unrelated parts, the parts within the formula are either related by simple conjunctive relationships (type III) or by synthetic relationships (type IV). Examples of these parts for type II could be: year of publication, name of magazine, country, language, author's last name, or principal subject, which are extracted from an article; headings of questionnaire data for a person; properties of substances from various sections. For type III, individual subjects, processes, conditions, methods, etc., for an ordinary sub ject heading could be extracted. Type IV can be illustrated by structural formulae of organic compounds.

With an insufficient degree of accuracy, such a subdivision can be further continued, provided that the first subdivision can be made on a one-to-one basis; in this process we shall come sooner or later to a small number of easily standardized characteristics. But if this subdivision, on the basis of meaning, on the basis of content of the e_k or the question, proves to be an impossible one or cannot be carried out on a one-to-one basis (or an almost one-to-one basis, since several different subdivisions are acceptable), then it is necessary, when approximating the total e_k or the question, or parts of them which were isolated as a result of the acceptable subdivisions, to resort to a more complicated method of indexing when referring to information not contained directly in the e_k. This indexing is actually a classification, that is, a referring of the entire e_k or the question, or parts of them, to one classification characteristic (heading)—or several independent ones—in a single system of standard characteristics of classification (type I), to one characteristic (or several independent ones) in each of several sections (Ranganathan facets) or classification systems (type II), or to several characteristics in one system jointly describing the content being indexed (type III).

Thus, there is a certain similarity between indexing and classifying (5). Indexing can be defined as direct, inductive classifying, which is more accurate than ordinary classifying, since it is also possible to take synthetic relations into consideration. Classifying can be viewed as less accurate, but a more nearly one-to-one, deductive indexing, where the role of the information language is played by the system of classification headings. Below we shall consider information language and types of information tasks from a more general point of view.

CONDITIONS FOR SELECTION

The conditions for selection evolve from the ordinary posing of the task of searching for information, when it is necessary to find—to collect—information devoted to each of the "subjects" making up the volume of the concept characterizing the question asked. This, by the way, is usually emphasized in the very way the questions are asked. For example, a person requests literature on the subject "properties of antibiotics," thus wishing to obtain information on the properties of penicillin, or streptomycin, etc., and does not formulate the question as "the properties of an antibiotic," which would mean the desire to find out information dealing only with properties common to all antibiotics. In other instances the way the task is posed cannot be reflected in ordinary language, when, for example, information is requested on the subject “the effect of radiant energy upon bacteria,” although even here it is understood that it would be more correct to say “various types of action of various types of radiation upon various types of bacteria.” With the aid of a stricter formulation of questions it is possible to reduce all the varieties of questions to such a generic-specific scheme. For example, when a person requests information on the “effect of acceleration upon an aircraft,” he usually wishes to obtain information about “types of action of accelerative forces upon various *parts of an aircraft,*” not upon various types of aircraft. Therefore, not “aircraft,” but “part of an aircraft” should be taken as one of the concepts characterizing the question.

Thus, the content of the e_k being sought as those dealing with the subjects which are part of the scope of the question-concept must be characterized by concepts whose content includes this question-concept. In this sense the logical function of the question, when satisfying the condition of selection, is the consequence of the logical function of the e_k. The conditions for selection will be described in more detail when the individual types of information tasks are examined.

INFORMATION LANGUAGE

The characteristics p_n, which form the universe *P,* and synthetic relations are analogous to the words and grammatical relations of ordinary language. Like the degree of approximation and the conditions for selection, they are determined by the nature of information contained in e_k. However, the appearance of new e_k may require the appearance of new characteristics to describe them, but the relations remain the same.

In order to observe the one-to-one correspondence of approximation it is desirable to have the characteristics (as well as the relations) not only standard, but also independent, that is, to assure that between them there do not exist any analytic relations which make it possible to substitute for some one characteristic another semantically similar characteristic or several characteristics which are united by means of the synthetic relations used in language. By analytic relations we understand here the constant semantic relations established between the characteristics on the force of a broader and older set of information than *E*. Obviously, the fewer relations that the language has, the more concrete and more varied the words must be to preserve the same degree of approximation, and, conversely, with the existence of many relations the number of characteristics and their concreteness can be fewer, and this makes it possible more easily to standardize them and ascertain their independence. The demand for independence of characteristics leads in the first instance to a com plicate representation of the less concrete, dependent characteristics in the form of disclosing their scope with the aid of enumerating a whole series of initial characteristics. In the second instance it leads to a complicated representation of very concrete characteristics in the form of compound complexes of a large number of initial characteristics (the content is disclosed). Ordinary language is rich both in relations and in words, and both are not independent, that is, use is made, for example, not only of a word designating a certain concept, but also words designating more general concepts (for example, generic for initial, which results in the non-one-to-one correspondence of representation of the content (without even considering synonyms), since instead of the first word it is possible to use its definition, consisting of other words which are related to one another. However, the presence of words for more general concepts does not make it necessary to describe them by means of a listing of words for the concepts making up their scope. Conversely, the presence of words for less general concepts does not make it necessary each time to give their complete definition by means of disclosing their content through words for more general concepts. Thus, the demand for the independence of characteristics leads to great difficulties in information language.

In order to surmount the difficulties mentioned, limitations are first imposed on the heterogeneity of content of the e_k themselves. One information task does not include e_k with contents which differ greatly with respect to degrees of concreteness. For example, searches for literature are not made simultaneously among books and articles. Analogous limitations are imposed on the questions asked, that is, it is required for the degree of concreteness of content of the questions to be within certain limits. Then choice is made of the most appropriate level of concreteness of the independent characteristics forming the universe *P*. With the aid of the generally used method of approximation, the characteristics from *P* must describe the content of e_k and the questions with an accuracy to any permissible degree of concreteness. If during this process the above-indicated difficulties are not eliminated then one selects two or more sets of independent characteristics *P*` *P*` *P*``, etc. without an intersection which correspond to various increasing levels of concreteness, and an approximation is made of the content of $e_k$ by means of the union of sets *P*=*P*` ∪*P*` ∪*P*``…, which thus consists of dependent characteristics. The content of the question is also approximated by means of a union of sets. In this process the above-mentioned difficulties are eliminated by using in the approximation of characteristics the most appropriate level of concreteness. That is, the particular content is described by characteristics which are as concrete as possible without using the listing of characteristics for disclosing the scope of the more general concept, if for some general concept this still proves impos sible even with the aid of characteristics at a lower level of concreteness, then it is considered to be beyond the limits of the accepted degree of concreteness of content and it is ignored, and, on the other hand, without the use of too complicated complexes of characteristics for revealing the content of a very concrete concept (if this proves impossible, the concept is replaced by a generic concept close to it). This method of approximation will be called the "natural" method.

Having avoided these difficulties, we nevertheless again come up against the possibility of the non-one-to-one representation of the content. Actually, the condition that characteristics which are as concrete as possible will be used requires the knowledge of all possible methods of representing them through less concrete characteristics, which (methods) can occur as a result of the dependence of the characteristics from P. On the other hand, the knowledge of these relations between characteristics is also necessary when searching for e_k on the basis of generic characteristics, when what is requested is not a listing of all concrete characteristics, but a description of a somewhat broader content with the aid of less concrete characteristics. Thus, the manifestation, in an information task, of a universe of characteristics P and analytic relations between them is closely related to the establishment of the method of approximating the content of e_k and the questions asked. This means that the development of a special information language is a necessary and, at the same time, the basic and most complicated, phase of making the process of task solution automatically controlled (1–3). This phase is obviously linked with works on terminology and classification and requires the knowledge of a broader scope of information than $E_,$ if the latter is not closed. The analytic links revealed between characteristics from P must be recorded in a form that makes it possible to locate them quickly and automatically when searching for e_k, and when approximating the contents of e_k and questions. The summarizing of these relations is similar to the set of cross references and other rules for the use of subject index. Thus, the locating of characteristics which are analytically related to one another constitutes an auxiliary information task (1), which, as a rule, deals with a broader content than E and which can be solved: *(a)* separately from the basic task when approximating the question asked, if only this does not result in a too complicated representation of the question, for example, in the form of listing a large number of characteristics revealing the volume of less concrete characteristics; *(b)* in the process of solution, when, for the content of the e_k being examined, more general representations of it than that which is recorded are found very quickly in a special memory device and all the representations are compared with the question, which is approximated, like the content of e_k in instance *(a),* by the natural method; *(c)* one time before recording the contents of the e_k with the subsequent recording of the result of solution of the auxiliary task when recording the content of the e_k itself, that is, by limiting ourselves to natural approximation for the question, we shall make for the e_k several approximations at various levels of concreteness, or, putting it more exactly, with a drop of one degree in the level of concreteness for each characteristic (a rise in the level of concreteness would, once again, require a complicated representation of the volume of less concrete characteristics, and a drop by several degrees of concreteness would again lead to a very complicated representation of very concrete characteristics, although sometimes this can be used with successful coding). For complicated information tasks with a broad volume of information from *E,* the first two methods of making the solution of the auxiliary information task automatically controlled are not very applicable, *(a)* because of the large amount of time expended when searching for e_k with a complicated and, as a rule, unnatural representation of the questions, and *(b)* because of the absence of a high-speed memory device of sufficient capacity for auxiliary searches. The third method has found applicability in cases when, with the aid of proper coding, the recording of approximations at various levels of concreteness can be made economic enough. Actually, the principal shortcoming of the third method of solution of the auxiliary information task is the necessity of recording each characteristic as many times as it is encountered, not only in its natural form, but also in the form of representation through characteristics of a lower level of concreteness, which reveal its content; that is, it is necessary to give its definition each time.

Let us note that the first two methods of solving the auxiliary information task presuppose that the transition to approximation of the content of e_k at a lower level of concreteness of characteristics can be made by means of representation of each individual characteristic through less concrete characteristics. In the third method of solution this cannot be presupposed; however, then this approximation will be made too labor-consuming and nonautomatic and will require a still greater increase in the number of entries. It will not be possible to reduce the number of entries by means of coding, since it will be necessary to give, for the entire content of the e_k, several parallel, individual approximations. Therefore we shall assume in the future that this transition is allowed by the method of approximation itself.

Almost everything that has been said with respect to characteristics is applicable to synthetic relations; however, taking into consideration the fact that the number of different relations in a language is usually not great and therefore the standardizing of them and the ascertaining of their dependence upon one another do not represent any great labor, we shall not dwell on them in greater detail.

Having thus ascertained that the consideration of analytic interdependencies among characteristics, as well as among synthetic relations, leads to an auxiliary information task that can have parallel solutions (separately or jointly), we can consider that the universe P consists of standard characteristics p_n which, when the basic task is solved separately from the auxiliary task, can be viewed as independent characteristics, with the approximation by them of the content of the e_k and the questions being carried out on a one-to-one basis—by the natural method. An analogous assumption can also be made with respect to synthetic relations. The types of information tasks revealed below by this assumption will obviously also occur for auxiliary information tasks, which, if they are viewed separately, do not in any way differ in principle from the basic tasks (some of the characteristics of the basic task begin to play the role of elements of information in the auxiliary task). The type of any information task will then be determined by the combination of types of the basic and auxiliary tasks into which the information task is broken down.

BASIC TYPES OF INFORMATION TASKS

TYPE I.

This is the simplest type of task, to which the other tasks can also be reduced, with certain tasks belonging exclusively to this type as a result of the insignificance of the information contained in the e_k. Here the content of each e_k is characterized simply by one or several standard names. These name-characteristics can be completely independent of one another. They can be given to elements of information to a considerable degree arbitrarily. Names which are similar (or even identical) in spelling can correspond to e_k which are completely dissimilar to one another. In instances when the e_k is characterized by several names, these latter are linked purely accidentally and other, different e_k can correspond to each of them.

The scope of the questions consists of the names encountered in all the e_k and also of other names which can be constructed analogously to the former. Having in mind, here and elsewhere in the future, only elementary questions which cannot be represented in the form of several questions which can be solved by independent, separate searches, it is possible to say that each question consists of one name. The conditions for selection lie in the requirement of complete correspondence between the question and the name of the e_k.

All semantic, classification relations are absent here. Examples of this could be: searches for telephone numbers, names of books, addresses, and the like, by persons' last names; translations or definitions of words by words; descriptions of magazines according to their standard names; and in general, code values by codes.

It is obvious that with complete independence of the name-characteristics, no auxiliary information searches are necessary and all that remains is the basic information task in its pure form.

Standard name-characteristics can also be nonindependent and can reflect the content of the e_k somewhat more thoroughly, that is, they can be given in a nonarbitrary manner. The solution of the basic task here remains the same as before. However, in this instance the purpose of the auxiliary information task will be to reveal the (analytic) relations between the characteristics, and, as has already been stated, this task can belong to any of the types introduced. These relations can also find their reflection in the very spelling of the characteristics, that is, characteristics which are similar in spelling can correspond to dependent characteristics and to e_k which are similar in content, and this can be used in coding to solve the auxiliary information task by the third method. This similarity is revealed partially in words characterizing similar concepts (searches for translations or definitions of words according to the words themselves). In every scientifically elaborated terminology and nomenclature, this similarity is reflected still more completely. Another example could be the searching for some substances according to the values of some one of their properties or, in general, the values of a continuous function according to the values of one argument.

TYPE II.

In the second type of information tasks each e_k is characterized not by one standard name, which cannot be the case, but more thoroughly by several sections—by one or several standard characteristics in each section, with several characteristics in one section being related purely accidentally, as formerly. Insofar as an elementary question, once again, contains here no more than one characteristic or its negation for each section, then, as in type I, any e_k can be subdivided into several e_k having no more than one characteristic in each section. The synthetic relations between characteristics are lacking here, if one does not include in them logical conjunctive and disjunctive relationships between predicate-characteristics and their negations. Actually, the content of any e_k here (after the subdivision into e_k containing no more than one characteristic in each section) is represented in the form of conjunction of the predicate-characteristic, and any nonelementary question in the form of a certain expression consisting of predicate-characteristics and their negations, united by conjunctive and disjunctive relations. The reduction of a nonelementary question to elementary ones is equivalent to the representation of that expression in the normal disjunctive form, so that each elementary question is represented in the form of the conjunction characteristics and their negations. In conformity with the general understanding of conditions of selection, it is required here that the conjunction representing an elementary question, or the expression representing a nonelementary question, be the logical consequence of the conjunction describing the element of information e_k which is selected out during the search. Since characteristics not included in the e_k can be introduced with a negation sign into a conjunction describing it, this demand can be reduced to having all the characteristics of the question and their negations contained among the characteristics and their negations from the element of information. That is, the set of the non-negated characteristics of the question must be included in the set e_k, but no negated characteristic of the question must be included in the e_k. This could be expressed in another form as follows. By assigning the symbol 1 to each characteristic which is contained in the e_k and the symbol 0 to each characteristic which is not contained in the e_k, and substituting these values into an expression for the question (not necessarily an elementary one) which is viewed as a Boolean function, we shall select the e_k only in the case that this function is reduced to 1.

Characteristics from various sections do not always depend (analytically) upon one another here. Within an individual section the characteristics may be not only dependent, but also independent. If the latter occurs for all sections, then obviously, auxiliary searches to determine analytically related characteristics will, as before, not be necessary.

Information tasks of the second type are a general instance with respect to tasks of the first type and are encountered much more often in practice. Only, as a result of the great complexity of such tasks, they have been reduced to tasks of the first type, when, for elements of information, we limited ourselves to the characteristics of the information contained in them only in one section, disregarding the other, occasionally no less important aspects. This corresponds to classification according to some one characteristic. However, as has been pointed out by many scientists, each event or substance is determined in the real, existing world by many coordinates, and a true classification must be multidimensional.

Searches for names and complete descriptions of substances, plants and bacteria, animals, and people according to the values of certain of their properties, names of diseases according to symptom complexes, bibliographical or even more complete descriptions of books, articles, and other publications according to the sets of characteristics from several sections (year, month, volume, country, language, author, type of publication, principal and secondary subject, and the like) and many other tasks (particularly statistical ones) pertain to this type.

They can briefly be characterized as the finding of values of functions of many independent variables.

TYPE III.

The third type of tasks differs from the second type of tasks only by the fact that it is impossible to record all the characteristics for a small number of sections (the total number of sections, as it were, greatly increases), in each of which it is possible to have just one or several completely unrelated characteristics. Therefore, without dwelling on such tasks in more detail, let us note, however, that the isolation of them into an independent type is linked with the appearance of a whole series of complications as compared with the second type of tasks, which complications, of both a logical and technical nature, will be noted below.

This type of task includes, for example, searches for names of diseases according to groups of symptoms, among which it is difficult to carry out a classification for a small number of groups of mutually exclusive (in the question) symptom-characteristics. Pain, fatigue, general debility, loss of weight, tendency to perspire, etc., can serve as examples of such symptoms. Here too one can include searches for complex concepts according to the simpler concepts comprising them, if one ignores the relations existing between the latter within the complex concept. Such searches can be encountered when preparing classifications, elaborating terminology, determining subject headings for subsequent basic searches of literature, etc. Let us note only that even here we assume that the characteristics have been properly standardized.

TYPE IV.

The fourth type of tasks differs from the second and third type by the fact that for elements of information and the questions asked, not only the characteristics are indicated, but also the synthetic relations between them within these elements and the questions. Once again, the relations, like the characteristics, are assumed to be strictly standardized.

The recording of the content of elements of information and the questions asked by means of interrelated characteristics is the most complete and most nearly perfect. The phrases of one language constitute one of the forms of this recording, although, true, a non-one-to-one and nonstandardized recording. Word order, expressions set off by commas, word combinations consisting of a modified word and modifying words agreeing with it in person, case, or tense, and other methods of recording the relations between the words of the sentence are the subject of a special science—grammar. It is also possible to cite examples where the characteristics are not necessarily situated consistently one after the other, but, as it were, correspond to points in space forming, together with the depictions of the relations, complicated spacial complexes. Such, for example, are structural formulae of organic compounds or geometric complexes themselves which characterize mathematical spaces.

A special variety of the fourth type of tasks is the search for structures where no distinction whatsoever is made between the characteristics in and of themselves, but consideration is made only of the relations among characteristics of the same type—points in space. If we do not make the distinction between relations, that is, if we consider that between each pair of characteristics, for example, there either exists a relation of a completely definite type or there does not, then we shall obtain an extreme instance, an example of which can be searched for linear complexes—graphs. Let us note that when searching for structures the distinction between characteristics and relations is beginning to become lost, and sometimes it proves more profitable to change over to a dual representation of mutually related characteristics, assuming the relations to be characteristics, and vice versa.

The questions in the fourth type of tasks consist, as a rule, of mutually related characteristics. A certain difficulty here lies in determining more accurately the conditions for selection. Apparently it is most natural, apart from the simple inclusion of the characteristics of the question or their negations among the characteristics or negations of the sought elements of information, to require that the relations existing in the question between the coincided characteristics of the question and the element of information be the consequence of the relations existing among the same characteristics in the e_k. Let us recall that we, as before, feel that with the existence of dependency among standard relations (as between characteristics), expanded conditions of coincidence are checked when solving the auxiliary information task.

By means of introducing new characteristics instead of relations, the fourth type of information task can sometimes be reduced to the second or third type of task. It is possible, first, in those instances when any relation between any characteristics can be replaced, without violating the accuracy of approximation, by assigning to each of these characteristics a so-called “role indicator” (3) describing the relation of the characteristic to the entire content of the e_k. For example, instead of an indication in a sentence with relations of the subject-object type with the aid of prepositions or case endings, it is possible to add the term “subject” to the word which is the subject, and to add the term “object” to the word which is the object, taking both words in the basic grammatical form (for example, for nouns, the nominative singular) and disregarding the prepositions. Introducing instead of each characteristic from P a group of new characteristics, each of which characteristics consists of the former characteristic and one of the "role indicators," we shall obtain a new set of characteristics, with the aid of which the content of e_k is approximated according to type II or III, but no longer with any indication of synthetic relations. The second instance when it is possible to eliminate the relations is the instance when the relation is established not between concrete characteristics, but between whole sections existing in the second type of tasks, ignoring the specific characteristics which exist in these sections for the e_k being examined. Such relations are actually new characteristics describing the entire content of the e_k. Putting it more exactly, each new characteristic is made up of the symbol for one of the former relations and the names of the sections united by means of that relation. Joining these new characteristics to *P,* we shall also reduce our task here to the second or third type.

Thus, in the future we can include in the fourth type only those information tasks which cannot be reduced, by means of changing the set of characteristics P, to type II or III tasks. Since in the fourth type of tasks the relations characterize not directly the e_k themselves, but are built up on the characteristics from *P,* it is no longer possible here, in contrast to types II and III tasks, to limit oneself to one-place predicate calculus, and in order to describe the logical nature of these tasks it is necessary to have a more complicated logical calculus. An example of the fourth type of task could be searches for structural formulae of organic compounds according to assigned fragments of those formulae, as well as, in general, searches for e_k with rather concrete content, when it is approximated by very general, independent characteristics. This situation requires the introduction of relations at least of the type of groupings in order first to describe smaller and less concrete parts of the content of the ek_by means of these characteristics, and then to describe the entire content of the e_k with the aid of groups of characteristics thus united. The latter task is encountered especially frequently when making a combined solution, by the third method, of a basic and auxiliary information task, when the dependent, more concrete characteristics are represented by less concrete characteristics.

CODING

CODE DESIGNATIONS AND SYMBOLS

Before describing some methods of solving information tasks of the types introduced above, we should like to discuss a few general considerations regarding the coding of standard characteristics and synthetic relations. This coding is necessary for making the process of solution—the comparison of the contents of the e_k and of the question—automatically controlled. In order to carry out the searches we must obviously record the approximations of contents of all e_k obtained during the process of indexing. This information must be recorded on some medium: in the lines of the subject index of the table of contents or table; in paragraphs of reference books, encyclopedias, etc.; on library catalogue cards; on punched cards, punched tape, magnetic tape, photographic film, etc. In addition to recording the approximation of the $e_{k,}$ it is sometimes possible to record the content itself, although frequently one limits himself merely to recording a standard name or number assigned to each e_k. According to the names or numbers found in the process of searches, one can then get to the content of the e_k itself, that is, once again the first type of information task is solved. Therefore, this method of describing the e_k must be avoided as much as possible if the task itself belongs to the first type. The approximation of the question must also be recorded somewhere: on a question list or in the memory of an information machine. Whereas during nonautomatically controlled searches the characteristics and synthetic relations are recorded in both approximations in the usual standard form—in the form of words, numbers, or other symbols—during automatic searches use is made not of this form, but of code designations consisting of groups of code symbols situated one way or another on the medium and in the memory. Code symbols can be, for example, notches or perforations on punched cards or punched tape, magnetized spots on magnetic tape, transparent squares on photographic film, and the like. In the memory these symbols can be represented by raised rods, closed contacts, etc. Ignoring for the time being in this discussion of coding the creation of code symbols, we shall consider that, in conformity with generally accepted considerations (according to the "yes or no" principle), use is made, both on the medium and in the memory, of only two possible states, one of which is assumed as the presence of the code symbol and can be designated by the figure 1, and the other as the absence of the code symbol and is then designated by 0. Thus, the code designations of the characteristics (and sometimes of the relations, about which more details will be given below) are represented in the form of binary numbers. In addition to the value of the number for the configuration of the code symbols within the field, various characteristics can also be distinguished by the location of that field on the carrier or in the memory set aside for the recording of the code designation.

CODING OF CHARACTERISTICS ACCORDING TO THEIR LOCATION (1, 2, 6)

The following types of codes are distinguished according to the location of the fields set aside for code designations of the individual characteristics:

1. Local code, when one knows beforehand the place where the field set

aside for recording some individual characteristic from *P* must be located; obviously, this is possible in the case of type I and type II tasks, assuming that e_k has just one standard name or no more than one characteristic in each section.

2. Direct code, which is a particular instance of a local code, when in each section it is possible to record only the presence or absence of a single characteristic. In this instance, the extent of the field is reduced to the minimum necessary for recording only either of the two possible code symbols 1 or 0. This code is applicable to type II tasks with an increase in the number of sections and a simultaneous decrease in the number of all possible characteristics capable of being found in each separate section to 1, or for the type III tasks, when the sections are absent but the total number of characteristics is not great and is equal to the number of the fields of the direct code.
3. Nonlocal code, when one does not know beforehand the location of the fields set aside for recording the individual characteristics, as is typical for type III and type IV tasks. The fields are situated consecutively so that they can be examined consecutively.
4. Superpositional code, which serves to eliminate the difficulties linked with the use of the nonlocal code, when all fields for the characteristics whose location cannot previously be anticipated are no longer situated one after the other, as in the case of the nonlocal code, but are superimposed on one another, that is, they coincide. This code can be used in the case of type III and type IV tasks only with certain limitations deduced with the aid of the theory of probabilities, since when fields are superimposed, there are formed parasitical configurations of symbols, which correspond to characteristics which are not actually included in the approximation of the e_k, and this leads to the selection of excessive, unnecessary e_k.

BASIC DEMANDS MADE UPON CODING

As we have seen, the use of certain of the above-enumerated codes is linked with the types of information tasks. In addition, it is also linked with one of the demands made upon coding—the demand of economy of recording. Actually, the establishment of a mutually one-to-one correspondence between sections and fields in the case of local code makes it possible not to record in the form of a configuration of symbols for the name of the sections themselves. In the case of direct code, this leads even to minimum configurations— to the use of just one of two symbols (0 or 1). On the other hand, in the case of the presence in the e_k of only an insignificant portion of the characteristics recorded by the direct code, we should have to reserve on the medium much additional place, and this is specifically precluded when converting to nonlocal code.

The second group of demands made upon coding evolves from the demand for simplicity in making automatically controlled the very process of comparing the code values of the characteristics from e_k and the question. From this point of view the most complicated is the nonlocal code, since, when it is used, it is necessary in general to compare each characteristic from the e_k with each characteristic of the question, whereas the knowledge of the place where the particular characteristic can be situated makes it possible to avoid this when using the other codes. Therefore, when using simpler devices for searches, the nonlocal code is replaced by the superpositional code. But this group of requirements also imposes limitations on the configurations of symbols within the fields of the local and nonlocal code (for the superpositional code there exist still greater limitations which are obtained, as has already been mentioned above, with the aid of the theory of probabilities, and which we shall not consider here). Assuming that code values of all characteristics recorded in some field of the local code are binary numbers with equal number of bits, this limitation can be formulated as follows. A binary number for any characteristic from one and the same field (section) must contain an identical number of 1's (1,2,6). This requirement is formulated analogously for all characteristics recorded by the nonlocal code. This code is called a “uniform” or “selector” code, as distinct from simply a binary code, when this limitation is absent.

The third group of demands made upon coding evolves from the demand of simplicity and automatic control in finding the code values themselves. In general the finding of the code designations of characteristics according to their standard characteristics represents a type I information task and can be arbitrarily called “tabular” coding. Therefore, for type I tasks a method of finding the code designations analogous to the recording of the content of the e_k on the medium in the form of a reference number, which was mentioned above, is not very applicable. An alternative to this method is the direct obtaining of the code designation with the aid of definite rules from the characteristic's standard name itself, which can be called “order” coding. Such, for example, is the usual order coding of the characteristic where the orders are understood as definite places upon which the standard symbols—letters, figures, and the like—stand in the characteristic. Since there usually are not many symbols used for recording a particular characteristic, it is easy to remember them or automatically obtain their code designation, from which a code designation of the entire characteristic is then made up. Before obtaining the code designations of the individual symbols, it is possible to have a standard conversion of the characteristic, for example, the reduction to a definite number of orders by means of the rejection of certain symbols, transposition and combination of symbols, etc. Standard conversion is also possible on the obtained code designations of symbols from the already converted characteristic. For example, a binary-decimal representation of numbers recorded in a decimal system of numeration can be converted into a purely binary representation, and to this it is possible to include the superimposing of code designations for individual symbols one upon the other, which is analogous to superpositional coding. All these conversions must be such as will preserve the mutually one-to-one correspondence between the characteristics and their code values and such that, at the same time, the number of orders in the code designations will be as small as possible. Here it is necessary to note that the third group of demands made upon coding lies, as in part also for the second group, in the contradiction with the demand of economy; and therefore in practice it is necessary to limit oneself to a compromise solution.

We spoke earlier only about coding the content of e_k, approximated by standard characteristics accepted when viewing the basic types of information tasks as independent ones. With the simultaneous solution of the basic and auxiliary information tasks by the third method, analogous demands also occur with respect to the coding of the dependent characteristics, since the less concrete characteristics approximate such more concrete characteristics for one of the four types considered above. With the aid of fortuitous coding it is sometimes possible to avoid a considerable increase in the volume of recording each dependent characteristic, if one reduces its former code designation to a combination of code designations approximating its less concrete characteristics.

CODING OF RELATIONS

Leaving to one side other, more detailed demands made upon coding, let us dwell in conclusion upon the coding of synthetic relations (1, 2). These relations are reflected onto the medium by the procedure of following the characteristics recorded by the nonlocal code and by the proper location alongside of the related characteristics of the code designations of those relations, each of which, once again, can be represented in the form of a binary number, or simply by a successive indication of the ordinal numerals of the related characteristics together with a code designation of this relation itself. The first method is similar to the local code for characteristics, and the second for nonlocal. In the memory the relations are reflected by the method of introducing into the comparison the question characteristics contained in it, by means of a particular cross-plugging of the detecting elements which detect the result of the comparison of each question characteristic with the characteristics of e_k, by the procedure in which these detecting elements must operate, and also, as for the relations recorded on the medium, by the proper location of the code designa tions of the relations in the memory. In view of the fact that the types of synthetic relations have not as yet been sufficiently ascertained, the more detailed consideration of the methods of coding them in abstracting from concrete information tasks is a difficult one.

SOME METHODS OF SOLVING INFORMATION TASKS

TYPE I TASKS

To solve such tasks, most often one uses ordinary dictionaries, reference books, encyclopedias, printed subject indexes (with standard subject headings), tables, card files, and the like. The characteristics $p_n \in P$ assigned to a particular $e_k \in E$ in its standard, uncoded form, are situated here in lexicographical order, and alongside of each characteristic one finds the answer $E_n \subset E$ consisting of all the e_k approximated by that characteristic. Therefore, in order to find for some question characteristic $\bar{p}_n$ the answer E_n, that is, the characteristic p_n coinciding with it, there is no need to examine successively the whole set of elements of information *E,* which is reduced here to a successive examination of the characteristics from *P,* but it is sufficient, for a small number of steps, to subdivide successively the entire *E* (or *P*) into as many subsets as there are different symbols which can be in the first order of the characteristic $\bar{p}_n$, and then, having selected the subset corresponding to the symbol contained in that order of $\bar{p}_n$, to subdivide it in conformity with symbols from the second order of $\bar{p}_n$, etc., until one finds the required answer in the form p_n, alongside of which the content of all the e_k from the subset E_n is directly recorded. In this process, in order to avoid duplication of descriptions of all e_k which are approximated by several p_n, each such description is placed only under one characteristic, to which references under other p_n are then given. In order to facilitate the process of subdividing *E* (or *P*) into subsets, the latter are placed in card files in different drawers or are subdivided by special inserts: similar arrangements are used in printed publications.

The described “devices” are the simplest parallel-action devices, since the question asked is simultaneously compared with many e_k, not one at a time, as occurs in consecutive-action devices. The former devices, when carrying out the search in a small number of steps, prove, as a rule, to have a much higher speed than the latter (even extremely improved ones), but are more cumbersome, since the comparison is made simultaneously with many e_k. Therefore, the simplest “devices” prove to be well adapted for solving type I tasks with large numbers of elements of information e_k (on the order of millions or hundreds of thousands), if it is not necessary for the time spent for the search to be very small and the searches themselves to be made automatically.

If such demands are present, one uses the most diverse varieties of parallel-and consecutive-action decoders. The characteristics p^n and $\bar{P}_n$ are represented in coded form, and, as has been mentioned already, in the event of type I tasks the coding must be done "by order" or otherwise we would again return to a type I task solved nonautomatically with the aid of the simplest "devices." For the same reason the answer must be given by the decoder in the form of the complete content of the information from $e_k \in E_n$ and not in the form of the intermediate names or numbers of those e_k, if only the number of all the $e_k \in E$ is not many times less than the number of all $p_n \in P$, which usually does not occur, since each e_k is approximated, as a rule, only by one p_n. Examples of such decoders could be, in the case of very small numbers of e_k (on the order of several tens), numerous varieties of mechanical and electrical decoders which very quickly give the value of the symbols encoded by their codes, and, in the case of small numbers of e_k (on the order of several thousands), various types of internal high-speed memory devices of modern numerical computers which produce on the "number lines" the content of the e_k almost immediately after the $\bar{P}_n$ is fed to the "address lines." For medium numbers of e_k (on the order of tens of thousands), the creation of devices which are just as high-speed proves to be extremely difficult, and, in addition, the applicability of such devices to solve general type I tasks is greatly limited by the small capacity of the memory cells set aside for the recording of the content of the e_k and by the necessity of spending time and foreseeing the method for extracting and decoding these records (recording of nonliteral and nonnumerical content here is almost completely precluded). Consecutive-action decoders, which use punched cards, punched tape, magnetic tape, photographic film, etc., are only partially free of the shortcomings mentioned and, although they do possess the proper automation of searches, in the case of medium numbers of e_k, despite the complexity of the decoders, they do not produce a reduction in the time of search as compared with the simplest "devices."

Mechanical and electromechanical parallel- or parallel-consecutive-action decoders prove to be better adapted for solving general type I tasks. These decoders are sufficiently automatic and possess, in the case of medium and large numbers of e_k, higher speed of searching than the above-mentioned consecutive-action decoders or the simplest "devices," and, in the case of small numbers of e_k, higher speed than the simplest "devices," but a somewhat slower speed than the high-speed memory devices of computers. Examples of such decoders could be the Amfis system proposed by Avakian (7) for large numbers of e_k and a simplified variety of the decoder, proposed in the next section, with superimposed punched cards for medium numbers of e_k. With the aid of these decoders the carrying out of that small number of steps into which the searches, in the case of the simplest "devices," are broken down, are made immediately automatic and are accelerated, and the frame of microfilm that is found as a result of the searches, with the content of $e_k \in E_n$ recorded in the usual form, appears on the screen of a microfilm viewer or on a television screen (for transmission to a distance). The use of microfilm to record the content of the e_k makes it possible, when using these rather simple decoders, to obtain, in addition to the automation and acceleration of searches, a considerable decrease in the volume of the entries themselves. Such devices can therefore be considered the ones with best prospects for the rapid and automatic solution of type I tasks and, in particular, for finding translations of words in other languages, for the rapid obtaining of information concerning addresses, telephone numbers, and the like, and for other similar tasks.

TYPE II TASKS

An estimate of the number of all possible questions which can be made up for all $e_k \in E$ in the case of type II tasks shows that this number usually proves to be much greater than the number of all possible e_k and p_n. Therefore, the use of the simplest "devices" in which all these questions are located in lexicographical order becomes here, as a rule, practically impossible as a result of the excessive increase in their volume.

When using parallel- and consecutive-action decoders, the characteristics can now be encoded not only by "order," but also by the "tabular" method, and it is possible to obtain as answers intermediate names or numbers of the e_k sought, according to which intermediate answers one can then, as in a type I task, make an additional search of the contents of the e_k themselves. In other words, in order to facilitate the solution of type II tasks, it is considered justified to reduce the individual stages to the solution of tasks of a simpler type, type I.

The process of solution with the aid of parallel-action decoders is broken up, as in the case of the simplest "devices," into a not very large number of steps, during which a search is made for all the smaller subsets from a set of elements of information E. However, because of the absence of the completely systematic location of the e_k (the impossibility of locating in lexicographical order all the questions pertaining to them), it is not possible here to limit oneself only to the consecutive isolation of subsets inserted in one another, but it is necessary also to search for the intersection of rather large subsets from E which correspond to the individual characteristics of the question. And this does not make it possible to use, in the case of type II tasks, certain of the simpler parallel-action decoders, as, for example, the Amfis system. Putting it more exactly, the process of searching can be described as the finding of the answer $E^{\star} \subset E$ in the form:

$$E^{\star} = E_{n1} \cap E_{n2} \cap \cdots E_{na},$$

where $E_{n1}, E_{n2}, \cdots, E_{na}$ are subsets from E which correspond to the characteristics of the question $\bar{p}_{n1}, \bar{p}_{n2}, \cdots, \bar{p}_{n\alpha}$, forming the set $P^{\star} \subset P$.

The subsets $E_{n1}, E_{n2}, \cdots, E_{na}$ can be easily found by the characteristics $\bar{p}_{n1}, \bar{p}_{n2}, \cdots, \bar{p}_{n\alpha}$ with the aid of the simplest "devices," as was done in type 1 tasks, if all the $\bar{p}_n \in P$ are situated in lexicographical order. If this situation is made for each section separately, then it is also possible to carry out the simultaneous search for all assigned characteristics. These subsets must, however, contain not the complete descriptions of the e_k included in them, but only the intermediate names or numbers of the latter, in order to avoid the duplication of those descriptions (each e_k is usually approximated by characteristics from all sections) and to facilitate the finding of the intersection of the subsets obtained. The nonautomatic finding of the intersection is used in Taube's coordinate indexing system (8) and is actually the sole possibility of using the simplest "devices" for solving type II tasks. Such a method of solution proves to be not only labor-consuming, but also requires a rather large amount of time, without even mentioning the necessity of solving several type I tasks (one task for each question characteristic, and one task for each finding of the complete descriptions of the e_k according to their intermediate names or numbers), and therefore it cannot be considered satisfactory.

Other parallel-action decoders used to solve type II tasks include punched cards with notches on the edges and superimposed punched cards (11) (aspect cards). These devices are to a certain degree dual one for another, as can be revealed from an examination of the principle of comparison of characteristics (1), which is common not only to them, but also to the majority of other decoders.

Let us assume that two characteristics are compared: $p_n \subset P_k$, where $P_k \subset P$ is a set of characteristics approximating the corresponding e_k, and $\bar{p}_n \in P^*$, and let us assume that the code designations having the identical number of bits, the designations of p_n on the medium and of $\bar{p}_n$ in the memory, contain an identical number of code symbols (represented in the form 1 on the medium and in the form $\bar{1}$ in the memory), that is, the code is a "selector" code. Then it is not difficult to show that the characteristics coincide in the instance, and only in the instance, that

$$M_0 \cap M_{\bar{1}} = 0 \qquad \text{or} \qquad M_1 \cap M_{\bar{0}} = 0,$$

where M_1 and M_0 are sets of bits of the code designation for p_n which contain 1 and 0 respectively, and $M_{\bar{1}}$ and $M_{\bar{0}}$ are analogous sets for $\bar{p}_n$. In this process, the bits of the medium and the memory which mutually correspond to one another in a one-to-one relation are considered to be identified. Depending upon how one realizes the pairs of states 0 and 1, and $\bar{0}$ and $\bar{1}$, and the signals θ and I characterizing respectively the absence or presence of intersection of sets M, one obtains different designs for decoders.

In the case of punched cards with notches on the edges, 1 is taken as the notched cell, $\bar{1}$ is a raised or inserted rod, and θ is the absence of meeting between the rod and the unnotched cell $(M_{\bar{1}} \cap M_0 = 0)$ causing the drop or falling out of punched cards corresponding to the sought e_k. Characteristics from P_k and $P^{\star}$ are encoded here according to local (in particular, direct) selector code, and the comparison can be done immediately for all characteristics from $P^{\star}$.

In the case of superimposed punched cards, 1 is taken as the punched-out cell, $\bar{1}$ is the superimposed card, and θ is the passage of light through the cell, with the absence of an obstacle in that cell on the superimposed punched cards $(M_{\bar{1}} \cap M_0 = 0)$, which passage of light detects the e_k corresponding to the open cells. Characteristics from P_k and $P\star$ are coded here usually only according to direct code, although later (1,9) it also proved possible to use a local selector code.

Thus, the pair consisting of rod (light) and moving punched card in the first instance realizes the pair $(\bar{1};\theta)$, and in the second instance the pair $(\theta;\bar{1})$, in which the above-mentioned duality of the devices considered manifests itself.

The devices described do not provide for complete automation of searches or for a considerable reduction in the time required for searching, as compared with the simplest "devices," although they are more effective than these latter. Therefore, the question arose concerning the development of more nearly perfect parallel-action decoders, and the use of consecutive-action decoders become justified (as distinct from type I tasks). More details about these decoders will be given in the next section, but here the only thing remaining is to recall that with medium numbers of e_k (on the order of tens of thousands) consecutive-action decoders do not have a sufficiently high speed of search. This speed, although it does exceed the speed of the devices just described, is still, even for extremely complicated decoders, on the same order as the speed of the simplest "devices" when solving type I tasks. For large numbers of e_k this ratio between the speeds of searching becomes still more unfavorable for consecutive-action decoders; their speed of searching must be almost the same, for example, as that of the last of the devices described in this section. Thus, the most suitable device for solving type II tasks must apparently be some kind of improved parallel-action decoder, which is somewhat more complicated than the Amfis system, but is simpler and faster than high-speed consecutive-action decoders. This finding is also supported by the fact that the locating of subsets $E_{n1}, E_{n2}, \cdots, E_{na}$ is reduced to the solving, by an already known method, of a type I tasks, and only the finding of the intersection of those sets can cause difficulties.

With small numbers of e_k it is possible to use as a more nearly perfect decoder various electric (1,10), electronic, and other decoders similar to inter nal high-speed parallel-action memory devices in present-day numerical computers. As distinct from type I tasks, the problem solved here is the extraction of the sought e_k not in the form of a complete description of the information contained in them, but in the form of intermediate number, which situation corresponds more to the possibilities (volume of cells in the memory) of those decoders, but, on the other hand, now it is necessary to have a separate output for each e_k, and it is impossible to get by with general output "number lines." Actually, whereas formerly one answer $E^\star$, consisting, as a rule, of a small number of elements of information e_k, was prepared for each question, and this answer was produced on the general output lines, now, as a result of the extremely large number of all possible questions, each e_k is recorded separately on the "address lines," and since all the e_k from $E^\star$ are searched simultaneously, it is no longer possible to lead them out onto the general output lines. In such decoders $\bar{0}$ can be taken as the presence of signals on the input "address lines," 1 as the presence of a connection of the output line corresponding to a separate e_k with an input line by means of a diode, capacitance, resistance, or the like, and θ is the absence of signals on the output line $(M_{\bar{0}} \cap M_1 = 0)$, which absence detects the necessary e_k (having taken as 1, and also for $\bar{1}$ the opposite state, we shall come to a still more obvious analogy with punched cards which are notched on the edges).

For medium numbers of e_k, the described decoders produced have as yet, unfortunately, been too complicated, and this makes it possible in this instance to attempt to improve immediately the system of superimposed punched cards, where the finding of the intersection of sets E_{n1}, E_{n2},· · ·, E_{na} is successfully made automatic, requiring, however, the recording of each set according to a direct code, that is, on each card (or set of cards) cells are set aside for all possible e_k. The essence of the improvement proposed lies first of all in the fact that the characteristics p_n are now recorded not according to the direct code, when each p_n has its own card, but according to a local selector code, which makes it possible to reduce sharply the total number of cards (1,9)—to reduce them to the number of bits in the code designation of the characteristics from all sections, and secondly, lies in the replacement of the superimposition of punched cards by a slight shift (1) of the always superimposed cards into an examination position similar to that used, for example, in the Card Translator. If each e_k is described by no more than one characteristic from each section, as occurs or can be achieved in type II tasks, then after the shifting of cards corresponding to the set of bits $M_{\bar{1}}$, for each characteristic of the question $\bar{p}_{n1}, \bar{p}_{n2}, \cdots, \bar{p}_{na}$ those, and only those, cells which correspond to the necessary e_k will allow light to pass through. Since the cards can be shifted automatically, for example, by means of electromagnets, the entire process of finding $E^\star$

proceeds extremely rapidly. With order encoding of characteristics from individual sections it is not necessary first to solve a type I tasks, since they are automatically solved on the decoder itself, and with the use of the position of each open cell on the surface of the card as a direct indication of the corresponding frame of microfilm it is possible automatically to receive on the screen a complete description of the e_k found.

If in the described device one limits himself only to one section, in which any characteristic is assigned completely, the latter can also be encoded by a binary, but not necessarily a selector, local code, which leads to a still greater decrease in the total number of cards, but which requires, however, two cells for the recording of each e_k. Such a simplified device can serve, as was mentioned above, for solving type I tasks. Let us note in conclusion that a still greater decrease in the number of cards can be achieved if one refrains from using just two code symbols; however, in this process, in order to record each e_k it is necessary to have a number of cells which is one greater than the number of code symbols, and, instead of the simple shifting of each card, it is necessary to shift it into one of several different positions, the number of which is equal to the number of code symbols.

For large numbers of e_k (on the order of millions or hundreds of thousands) the use of the device described proves to be a difficult one, since, in order to set aside on the card a cell for each e_k, it is necessary to use cards which are too large (the size of the cell is limited from the bottom, in order to provide for the unhindered passage of light through all the punched cards). The replacement of each card by a series of cards leads to a manyfold duplication of the entire device. The way out of this position must apparently be sought (if one does not return again to complex nonmechanical decoders) in rejecting the direct code for recording sets $E_{n1}, E_{n2}, \cdots, E_{na}$, and in recording the answer E_n for each characteristic p_n according to a nonlocal code on a group of punched cards or on a piece of punched tape, photographic film, or the like, in the form of a sequence of code designations of numbers of all $e_k \in E_n$. Having found consistently the E_n corresponding to the individual characteristics of the question, by means of the simplest "devices" or a device of the type of the Amfis system we isolate the E_n which contains the smallest number of e_k. Then the numbers e_k from that E_n are introduced into the memory device of a consecutive-action decoder, and the content of each other found E_n is consecutively examined by that decoder. Those numbers in the memory of the decoder for which the coincidence in examination of each E_n will be detected will correspond to the sought e_k forming the intersection of all the located E_n. Such a parallel-consecutive method of searches will obviously prove more profitable than the purely consecutive method if the number of e_k from all the examined E_n, multiplied by the quotient obtained when the number of e_k from the E_n placed in the memory is divided by the number of numerical designations in that device, is much less than the total number of $e_k \in E$. With a large number of characteristics p_n in each section and a not very large number of the sections themselves, such a situation ordinarily occurs. Actually, considering that the characteristics in each section were chosen to be as homogeneous as possible in their concreteness and that approximately identical numbers of e_k correspond to each of them, and remembering that each e_k is described by no more than one characteristic in each section, we can approximately evaluate the capacity of the E_n as the quotient obtained when the number of $e_k \in E$ is divided by the number p_n from the corresponding section. Obviously, the highest degree of automation will be achieved in such a device if the input of E_n into the memory and the examination of the remaining E_n are carried out directly from the screen of a system on the type of Amfis, for example, with the aid of an iconoscope.

In conclusion let us note that with the presence of a negated characteristic in the question, the methods of solutions remain approximately the same, except that the e_k found for the negated characteristic, if it is viewed as nonnegated, are not left in the intersection of the E_n corresponding to the nonnegated characteristics, but, on the contrary, are rejected. Sometimes such e_k are revealed when this intersection is compared with a new intersection obtained when a negated characteristic viewed as a nonnegated one is added.

TYPE III TASKS

In order to solve these tasks one may use the parallel-action decoders described in the preceding section, only in the instance that the characteristics of p_n from the set P_k which describes the corresponding element of information e_k, are recorded by direct code, that is, each $p_n \in P$ is given its own place in the simplest "devices," its separate card in a system of superimposed punched cards, its cell on punched cards with notches on the edges, and its input line in electrical decoders. The attempt to use a local code to record the characteristics would lead to a situation in which, not knowing in what field a particular question characteristic could be found, we would have to ask the question by very many methods, the number of which, at best (with a systematized sequence of characteristics), is equal to the number of all possible combinations from among the fields of the local code for the number of question characteristics. With, as a rule the large number of all possible p_n the last two decoders just listed cannot be considered practically feasible. Thus, the only methods remaining are the first two, relatively nonautomatic methods, and of the improved decoders, only the last of those described in the preceding section.

In order to make use of the remaining parallel-action decoders it is necessary to record the characteristics from P_k according to a superpositional code. As was already mentioned earlier, it is possible during this process, however, for superfluous, unnecessary e_k to appear during the searches. Without considering here that possibility, which has been worked out in detail by many researchers, let us change over to a brief description of consecutive-action decoders. Although, in conformity with what has previously been set forth, they can be considered to be satisfactory only for medium numbers of e_k, as a result of the absence at the present time of the last of the improved parallel-action decoders described in the previous section, they are the only devices for the automatic solution of type III tasks (without the use of superpositional code).

The basic shortcoming of a consecutive-action decoder is the necessity of a very rapid, consecutive examination of the content of each e_k. Sectors of the medium which contain the code designations of the characteristics from P_k which are recorded by a nonlocal code are consecutively passed or switched up to the scanning device of the decoder. Then each p_n from the P_k is compared with all the $\bar{p}_n$ from the set of question characteristics $P^\star$. If the $P^\star$ is included in the P_k ($P^\star \subset P_k$), the corresponding *ek* is selected.

If the medium is a discrete one (punched cards, cards made out of film, or the like), it is possible to make a consecutive comparison first with one $\bar{p}_n \in P^\star$, then to compare the e_k selected with another $\bar{p}_n \in P^\star$, etc. Taking into consideration the fact, which was already mentioned above, that usually only a small number of all the $e_k \in E$ proves to be approximated by a certain characteristic, it may be considered that the number of repeatedly examined e_k will decrease rapidly, that is, it is completely admissible, with a discrete medium, to use the simplest decoders with a memory only for one question characteristic in order to solve type III tasks (an example could be the Samain (12) selector—the Filmorex system).

For a continuous medium it would be necessary by such a method to rerecord the content of the e_k selected during each examination. Therefore, the comparison of each $p_n \in P_k$ is carried out immediately with all $\bar{p}_n \in P^\star$. The decoder required for such a comparison is analogous to the electric, electronic, or other decoder described in the preceding section, in which decoder the role of the question characteristic $\bar{p}_n$ is played by an individual characteristic $p_n \in P_k$, and the role of the code designations of the p_n is played by the code designations of all $\bar{p}_n \in P^\star$ which are recorded on the input lines with the aid of diodes, resistances, etc. Thus for 0 one can here take the presence of a signal on the input line, for $\bar{1}$ the presence of connection of the output line of the corresponding individual $\bar{p}_n$ to the input line with the aid of a diode, resistan*ce,* etc. (or directly in the case of a single $\bar{p}_n$ in the question), and θ is the absence of a signal on the output line $(M_0 \cap M_{\bar{1}} = 0)$, which absence detects the coincidence of the corresponding $\bar{p}_n$ with the p_n being examined. If, after the examination of a certain e_k, the coincidence will be recorded for all $\bar{p}_n \in P^{\star}$, then this e_k is selected or its content is re-recorded. Since the number of in the question is usually not large, the decoder proves to be extremely simple, although it does require a special detecting element for each $\bar{p}_n$. The outputs of the detecting elements can be joined by means of switches in such a way as to detect not only the value of the conjunction of the nonnegated characteristics, but also the value of an arbitrary Boolean function of the question characteristics (1,2,6,13,14), that is, the e_k can be selected during one examination even for a nonelementary question (in particular it is possible to search simultaneously for many questions if the decoder has a sufficiently large memory).

Depending upon the type of medium, the signals 0 can be fed to the input lines of the decoder or directly from the brushes through the holes in the punched card or punched tape, or, after amplification, from the heads picking the magnetized spots of the magnetic tape or drum, or from photocells reacting to the transparent spots of photographic film or photodisc, etc. Other designs of decoders are also possible, provided they are of such a high speed that they do not limit the speed of examination of the e_k. Since they contain a small number of $\bar{p}_n$, a certain complication of them (both conditions, $M_0 \cap M_{\bar{1}} = 0$ and $M_1 \cap M_{\bar{0}} = 0$, are checked) is also possible, and this complication makes it possible to record characteristics p_n and p_n according to a binary, but not selector, code. The medium with the best prospects is apparently photographic film (frames of film, photodiscs), not only as a result of its high degree of resolution, but also because of the possibility of recording and extracting complete descriptions of e_k, instead of just intermediate numbers, which would require an additional solution of a type I task.

Thus, with medium numbers of e_k the devices that can be considered the best adapted for the automatic solution of type III tasks are consecutive-action decoders or an improved decoder with superimposed punched cards and with the use of a superpositional code, but with large numbers of e_k the best device is the improved parallel-consecutive-action device which was described at the end of the preceding section.

TYPE IV TASKS

Generally speaking, only consecutive-action decoders can be employed for the solution of these tasks, since the synthetic relations do not characterize directly the e_k themselves, but are built up on characteristics from e_k. Parallel-action decoders could be employed here only for the preliminary rough selection of e_k according to $\bar{p}_n$ which are assigned, as in a type III task, without any indication of the relations. Although the number of e_k found during this process proves to be less than the number of all $e_k \in E$, nevertheless it is much greater than the number of the e_k sought, or otherwise there would be no reason to show the relations at all. For medium numbers of e_k the automatic solution of a type III task (with the aid of a parallel-action device) produces, as we have seen, only the numbers of these e_k, and therefore the locating by numerical designations of a rather large number of complete descriptions of these e_k, as in type I tasks, can occupy too much time for subsequent searches taking the relations into consideration.

Consecutive-action decoders used to solve type IV tasks differ from those mentioned in the preceding section only by the presence in the memory of additional detecting elements, switches, and counters for comparing the relations. Taking into consideration the certain lack of definiteness and lack of study of this type of tasks, it is for the time being difficult to give a general description of the requirements made upon the composition of such memory elements. For example, with very complicated relations it may be necessary to have a whole arsenal of facilities existing in modern numerical computers (15). With the existence of only the simplest relations of the grouping type (the uniting of some characteristics from P_k into groups which are then viewed as independent characteristics), it is sufficient to combine the outputs of the detecting elements corresponding to the characteristics grouped in the question, through the switch "and" and then to feed the output from this switch to an additional element which detects the inclusion of that group into a grouping of characteristics from P_k after the examination of that latter grouping, which situation is signalized by means of special symbols, "brackets"— the code symbols are recorded on the medium in definite places at the beginning and end of the grouping. Devices possessing such possibilities are, for example, the Experimental Information Machine (1,2,6) of the Institute of Scientific Information, Academy of Sciences USSR (EIM), the WRU selector (13), and the planned Kodak Minicard system (14).

In the event that the characteristics from P are too general, as can occur with the presence of a complicated system of relations, it sometimes proves possible to have very many methods of establishing the mutually one-to-one correspondence between $\bar{p}_n \in P^\star$ and the $p_n \in P_k$ coinciding with them. Since only for certain of these methods can the relations between $\bar{p}_n$ be generated by relations existing between the characteristics from P_k which correspond to those $\bar{p}_n$, it is necessary to analyze all the possible instances of establishment of this mutual one-to-one correspondence. For this purpose it may be necessary to have a repeated reproduction of the content of each e_k. In order to be spared the repeated passage of a moving medium, in this instance it is necessary to rerecord the content of the e_k on an extremely high-speed auxiliary operational memory (16), and also to set aside a portion of this device for recording the intermediate results of the analysis and the program of its execution. Thus, the solution of type IV tasks in general requires the application of modern universal or even special digital computers. It must, however, be noted that at the present time there exist only a very few information tasks in which the approximation of e_k and the questions could be carried out on a one-to-one basis with such great accuracy and complexity as to require the solution of those tasks by means of such highly improved devices.

REFERENCES

1. V.P.CHERENIN, *Nekotorye problemy dokumentatsii i mekhanizatsiya informatsionnykh poiskov* [Certain Problems of Documentation and Mechanization of Information Searches], Moscow, 1955.
2. B.M.RAKOV, V.P.CHERENIN, *Eksperimental'naya informatsionnaya mashina Instituta nauchnoy informatsii AN SSSR* [Experimental Information Machine of the Institute of Scientific Information, Academy of Sciences USSR], Moscow, 1955 (there also is a publication in the English language).
3. J.W.PERRY, A.KENT, M.M.BERRY, *Machine Literature Searching,* Interscience, New York-London, 1956.
4. W.C.ADAIR, Citation Indexes for Scientific Literature? *American Documentation, 6,* [1], 1955.
5. B.C.VICKERY, Developments in Subject Indexing, *The Journal of Documentation, 11,* [1], 1955.
6. B.M.RAKOV, V.P.CHERENIN, *Byulleten YUNESKO dlya bibliotek;* Machines for Retrieving Information in the USSR, *Unesco Library Bulletin, 11,* [8–9], 1957. (Also published in English, French, and Spanish languages; the article was reprinted in the German language in the journal *Nachrichten fur Dokumentation, 9* [1], 1958.
7. E.A.AVAKIAN, E.GARFIELD, AMFIS—The Automatic Microfilm Information System, *Special Libraries,* April, 1957.
8. M.TAUBE and ASSOCIATES, *Studies in Coordinate Indexing,* Vol. I, 1953.
9. Same, Vol. II, 1954.
10. Same, Vol. III, 1956.
11. R.S.CASEY, J.W.PERRY, *Punched Cards—Their Applications to Science and Industry,* Reinhold, New York, 1951.
12. J.SAMAIN, Filmorex—*Une nouvelle technique de classement et de selection des documents et des informations,* Paris, 1952.
13. J.W.PERRY and ALLEN KENT, Das lesende Selektiongerät der Western Reserve University, *Nachrichten für Dokumentation,* No. 2, 1957.
14. A.W.TYLER, W.L.MYERS, J.W.KUIPERS, The Application of the Kodak Minicard System to Problems of Documentation, *American Documentation, 6,* [1], 1955.
15. A.OPLER, T.R.NORTON, New Speed to Structural Searches, *Chemical and Engineering News,* June 4, 1956.
16. W.H.T.DAVISON and M.GORDON, Sorting for Chemical Groups Using Gordon-Kendall-Davison Ciphers, *American Documentation,* 8, [3], 1957.

Subject Analysis for Information Retrieval

B.C.VICKERY

In another paper presented to this Conference "The structure of information retrieval systems" (Area 6), I have described some syntactic aspects of a retrieval system as a *lattice of units of information*. The elements of the lattice are *terms,* and these are linked in a network of *interlocking, inclusion,* and *coordinate* relations. In searching for a particular unit of information, the system can be designed to retrieve not only items recorded for the named subject of search, but also items recorded for subjects which *(a)* include, *(b)* are included by, or *(c)* are coordinate with that subject, since these related subjects may be relevant. The *limits of relevance* can be varied at the discretion of the designer of the retrieval system.

In the present paper I wish to consider some of the semantic and pragmatic problems encountered in constructing an information lattice and in defining the optimum relevance limits. In any retrieval system, the subjects indexed and sought are expressed initially in *words*. Relations between words must be considered in designing the system, at two stages: (1) in choosing what words are to be used as indexing terms (descriptors, index sets), and (2) in deciding what related terms (if any) are to be retrieved when a particular term is sought. The second stage can only be ignored by a system which relies solely on independent descriptors, used either in correlation or in simple dictionary form. The first stage cannot be avoided by any system. Relation between terms is the central semantic problem of subject indexing.

METHODS OF ANALYSIS

The first and most elementary of semantic problems in the design of an index is to equate synonyms, and this involves establishing *inclusion* relations between words: if word A includes B, and B includes A, then A and B are synonymous. Thus the chemical terms "ethanol" and "ethyl alcohol" are synonyms. But

B.C.VICKERY Imperial Chemical Industries Ltd., Welwyn, England.

the interrelating of words reveals another type of synonym, in which A is a phrase while B is a single word: thus nephrosclerosis is a synonym for hardening of the kidney. The elimination of such synonyms demands the analysis of a word such as nephrosclerosis into a combination of others at a lower *semantic level*.

This general phenomenon raises a number of problems. What modes of semantic analysis are available? In what ways can one term be part of another? How far should we proceed with such analysis? At what semantic level should we stop? Various answers have been given to these questions by those working on subject indexing, classification, and mechanical selection.

Three modes of analysis are distinguished in traditional logic: the *physical analysis* of a thing into its parts or constituents, or of a group into its members; the *logical analysis* of a generic concept into its species; and the *metaphysical analysis* of a concept into its attributes. A special instance of the last is the analysis of the *definition* of a concept into its elements. All these forms of analysis are used in different systems of information retrieval.

For example, it is general in the indexing of chemical substances to replace the trivial name of a chemical, a single word, by a compound term derived by physical analysis: the parts used are either functional groups (in standard nomenclature and in recent “ciphers”) or chemical elements (in formula indexes). The replacement of a single term by its constituent species is perhaps rare, but the use of logical analysis to produce a hierarchical classification of terms is well known. The representation of a concept by a combination of attributes is found in a number of correlative indexes for botanical identification, e.g., a particular fungus, *Amanita muscaria,* is represented by Findlay as a compound of the following indexing terms: Pileus large, flat smooth, orange, soft, Flesh thick, white, Spores colourless, moderate, elliptical, Stalk white, central, long, fleshy, Gills thick, white. The use of definition in subject analysis has been developed by J.W.Perry and his associates, thus a Thermometer is represented as “Device for measuring temperature.” Andrews and Newman at the Office of Research and Development, U.S. Patent Office, represent a word by the combination of a limited number of attributes, e.g., a Pitcher (from which a Measure is filled) is represented as an “Apparatus for containing and dispensing, with a lip”—what might be called an “operational definition.”

Similar techniques of semantic analysis have been used by some planners of auxiliary languages. For example, John Wilkins in 1668 represented Counter-poison (i.e., Antidote) as “Medicine against poison,” and Medicine itself as a compound, “Medicating thing.” In 1942, Lancelot Hogben formed compounds such as “agricultural person” for Farmer and “book-house” for Library. Both these authors used the technique of analysis by definition.

Any method, of semantic analysis must, however, face the problem of deciding when to stop. It is clear that there is no ultimate elementary level at which it should cease. The level of semantic analysis must depend on the purpose for which it is undertaken, and must be defined in some way by the designer of the indexing system. Techniques for defining the required semantic level are therefore needed, and have been provided by workers in this field.

CHOICE OF INDEXING TERMS

Let us consider first the method of "analysis by definition" used by Perry *et al.* They have analysed about 30,000 words in science and technology, and claim that a few hundred "semantic factors" (i.e., terms not further analysed) are sufficient to represent them. Examples of their analysis are:

abattoir=factory−destroy−animals

abortifacient=drug−destroy−embryos

induction furnace=machine−electricity−heat

Analysis by definition could of course proceed further. Thus, using *Webster's Dictionary* as did Perry, we might make such analyses as:

drug=substance−medicine

factory=business−building

On the other hand, dictionary definitions of other terms provide only synonyms, e.g.,

destroy=undo, ruin, demolish, etc., etc.

or are purely descriptive, e.g.,

heat=that which causes a rise in temperature.

In these last instances, analysis by definition is halted at the level "destroy" or "heat".

In other cases, some criterion must be laid down as to the level at which analysis is to be stopped. A full explanation has not yet been provided by Perry (his *Semantic Code Dictionary* is not available at the time of writing) but in *Machine Literature Searching* he and his colleagues related that "it was evident from the start that the analysis of terminology would be facilitated by considering groups of related terms". They give citations to psychological techniques of forming general concepts, and described their own formulation of "five very general classes": processes; machines, apparatus, devices; materials, substances; attributes, characteristics; and abstract concepts. It appears, therefore, that when analysis by definition arrived at semantic factors which could be allocated to one of these (and perhaps other) general classes, analysis was halted at that level. Whether this interpretation is correct or not, it is worth pointing out that the choice of level of analysis is here aided, if not controlled, by establishing a number of general classes or categories.

Let us turn now to the "analysis by operational definition" of Andrews and Newman. They pointed out that "the great bulk of things which we refer to are given functional names because of the process they perform…or the use to which they are put…. Names and other words used as descriptors usually infer either a broad relationship with some other unidentified thing, or an indefinite or undefined relationship with a specific thing." Andrews and Newman provided a series of "modulants," e.g., process, apparatus, work (product, starting material, intermediate), condition, made-from, and combination-including. They then chose "ruly roots" which, inflected by the modulants, formed their descriptors. The process of analysis used is clearly the opposite of this: in order to extract "ruly roots" from the named-things provided by the literature, i.e., in order to control the semantic level of these roots, Andrews and Newman found it helpful to formulate a series of modulants, once again, a series of categories.

Thirdly, let us consider the technique of classification known as "facet analysis," by Ranganathan. It consists of taking each of the terms used in a given subject field and defining it with respect to its parent class. Thus in the field of chemistry, "alcohol" is a kind of chemical substance, "liquid" a state of that substance, "volatility" a property, and "combustion" a reaction of it, "analysis" an operation on it, and "burette" a device for performing an operation. Having defined terms in this way, facet analysis sorts them out into the categories so formed, substance, state, property, reaction, operation, device, so that the categories can be combined together to form compound terms.

CATEGORIES

In all three techniques of analysis, therefore, the choice of semantic level of indexing terms is aided or controlled by the formulation of *categories:* "concepts of high generality and wide application, fabricated by the mind with direct or indirect reference to the experiential world, and employed by the mind in the interpretation of that world." It was no doubt recognition of this common feature of several techniques of analysis which led the International Study Conference on Classification for Information Retrieval (Dorking, May 1957) to conclude that "there is general agreement that the most helpful form of classification scheme for information retrieval is one which groups terms into well-defined categories."

It is further of interest and importance to note that the three techniques of analysis considered above, applied in the field of science and technology, have each isolated very similar lists of categories for information retrieval. Thus we may compare (A) Perry's "analytic relations," which are an extension of his "general classes," (B) Andrews and Newman's "modulants," and (C) Vickery's "facets," as follows:

A	*B*	*C*
Class inclusion	Process	Substance, product
Material of composition	Apparatus	Organ or part
Whole-part	Product	Constituent
Process	Starting-material	Property
Agent	Intermediate	Patient
Patient or product	Condition	Action, operation or process
Attributive	Made-from	Agent
Negative	Combination-including	Apparatus

If we roughly equate Attributive, Condition, and Property, then the only categories not present in all three lists are Class inclusion in A (and this is in fact dealt with in other ways by B and C), Negative in A and Intermediate in B. It appears therefore, that the control of the semantic level of indexing terms by categories is leading to similar results in these three techniques of analysis. Although there is as yet no agreed standard level of semantic analysis for indexing terms, yet an examination of categories used in different forms of retrieval systems does suggest that a considerable degree of uniformity is present.

RELATIONS BETWEEN TERMS IN COMBINATION

Most modern retrieval systems do not use indexing terms in isolation: they combine two or more terms in one way or another (the technique has many names: combinatory, coordinate, correlative, associative, multi-aspect, analytico-synthetic, and so on). At this stage, too, various levels of analysis are possible.

Indexing terms can be combined by simple juxtaposition, as in the systems of superimposable aspect or peephole cards, or in punched cards with superimposed fields, or again in such classification schemes as the U.D.C. where terms may be linked together by a semantically empty "fence," the colon. In such systems, all possible relations between terms in compounds are treated as identical. There is no discrimination into more specific relations.

The need to specify relations arises in two ways. First, a given combination of indexing terms may have more than one meaning; the compound Bacteria-Destruction-Dyestuffs may mean the destruction of bacteria by dyestuffs, or of dyestuffs by bacteria. To discriminate between these subjects, prepositional re lations are needed. Second, if the subject is a correlation between Cat-Feeding and Bee-Population, simple compounding could also imply a correlation between Cat-Population and Bee-Feeding. Some method of showing an interlocking relation within each pair of terms is needed.

Various techniques have been used to introduce specific relations between terms in a compound. In some systems, such relations are implicit in the *categorisation* of indexing terms. Thus the complex word "haemometer" may be analysed into the following indexing terms, each of which is allocated to a specific category (in parentheses):

Blood (substance)−flow (process)−rate (property of process)−measurement (operation on property)−instrument (device for operation)

If the categories are clearly distinguished as above, the *interlocking* relations between successive terms in the compound are made apparent. The combination of two categories, e.g., Operation and Device, implies a specific relation between the terms in each category. Therefore the use of categories leads to specifying relations between terms in a compound. The "analytic relations" between semantic factors and the word that is factored, the "modulant" relations between "ruly roots" and the named-thing that is analysed, and the relations between facets and the field that is analysed—all these imply relations within a compound *between* factors, modulants, or facets.

The next level of analysis of relations between terms in a compound is to provide further specifically identified *relational particles* to link related terms. Farradane, indeed, relies wholly on nine such particles, the "operators," and does not categorise the indexing terms. The faceted classification scheme of Ranganathan introduces six specific "phase" relations. Perry and his colleagues have isolated at least twelve "synthetic relations" between terms in a compound (starting material, material processed, containing, properties for, properties of, process, means, condition or circumstance, discussion of, location, attributive, field).

A deeper level of analysis of relations between terms in a compound has been suggested by Andrews and Newman, who give as examples of "interrelational concepts" Cause, How, Means, Thru, and a number of highly specific temporal relations. In briefly discussing their work, Vickery has suggested a number of other logical, spatial, and spatio-temporal relations which may be useful.

As regards the level of analysis of relations within a compound, therefore, current practices (or current proposals) are far more widely varied than in the choice of indexing terms. It seems certain that this variation is due to the varied purposes for which different systems are designed. The more specific the subject matter to be indexed, and the greater the volume of items to be specified and selected, so much the greater is the need felt for detailed analysis of relations between terms in a compound.

The problem for information retrieval in this region is therefore complex. First, it is necessary to continue the work of identifying relations between terms which will aid selection. Second, it is necessary to establish what levels of analysis of relations are useful in different retrieval situations, since it is wasteful to build into a system more discrimination than is necessary to select a given document. Third, it is necessary to design systems so that increasing discrimination—increasingly fine analysis of relations—can be smoothly fed in as the volume and specificity of the system increases.

RELATIONS WITHIN EACH CATEGORY

The pattern of the information lattice which emerges from the preceding discussion is an assembly of indexing terms (descriptors, index sets) sorted into categories, and a variable number of relational particles which may be used to link terms in a compound. The relation of a category to the subject field, of a category to other categories, of a term to its compound, and of a term to other terms in a compound—these do not exhaust the possible relations between words which are of interest and value in subject indexing. We have also to consider the relations between terms *within* a category.

Systems have been discussed, e.g., by Luhn, in which these relations are not specified. The descriptor (notion) used in the indexing system is not the individual term within a category, but the category itself. All terms in a given category are treated as equivalent. At the opposite extreme we have the typical faceted classification scheme, in which the terms in each category are arranged in a *hierarchy* of subordinate and coordinate relations, and the descriptor (class number) is a symbol which expresses the exact position of the term in the hierarchy, i.e., its relations to adjacent terms in the hierarchy. An intermediate solution is to list and use all terms in each category, but not to express hierarchical relations between them.

The first solution (category descriptors) assumes that the likeness between terms in a category is so great, and the unlikeness so small, that it is advantageous to retrieve all of them if any one of them is sought. The third solution (random code descriptors) assumes the reverse—that the unlikeness is so great, and the likeness so small, that there is no advantage in retrieving any other term in a category if a particular one is sought. The second solution (hierarchical descriptors) tries to arrange the terms in a category according to their degree of likeness and unlikeness, and offers the possibility of prescribing for each particular search what degree of likeness to the sought term is relevant.

Hierarchical arrangement in a category raises several problems. In the first place, it is a fairly frequent occurrence that a given term may belong to more than one helpful inclusion chain, thus Eggs, for example, may figure in hierarchies relating to Ornithology, Poultry, Nutrition, Cookery, Food hygiene, Folklore, etc. If a given retrieval system aims to serve users in all these fields, then provision must be made for all the inclusion chains, and for the selection of a given chain according to the interests of the searcher. As the aforementioned International Study Conference on Classification concluded, "in constructing schemes of classification and in applying them to a retrieval system, the fullest consideration must be given to providing alternative approaches for different users. In particular, freedom to vary the manner of combining categories and to vary the arrangement of terms in a category in different contexts must be provided."

The second problem of hierarchical arrangement is that the terms allocated to a category may fall into more than one hierarchy. For example, in the category "Soils" we may have five independent hierarchies: soils classified according to constitution, origin, physiography, texture, and climate. In other cases the various hierarchies are not independent. The category may be subjected to "metaphysical" analysis into its attributes, and a given term in the category may be formed by compounding attributes. The representation in this way of a fungus, *Amanita,* has already been mentioned. The representation of a chemical substance as a compound of functional groups is the rule in chemical nomenclature and modern coding systems.

Finally we have the problem of coordinate relations in a hierarchy, i.e., relations between terms all of which are subordinate to the same inclusive term. To form any such links in an information lattice might be regarded as illogical. The possible relevance to a given subject of documentary items whose subject includes, or is included by, the sought subject is clear but, in traditional classification theory, *coordinate* terms should be mutually exclusive. However, even if this is true of the terms, it is not true of coordinate subjects, and still less true of the documents. Consequently there is value in arranging coordinate terms in a series which brings closest together those most alike, and separates those most unlike.

There is no doubt that a retrieval system which incorporates subordinate and coordinate relations in each category is much more flexible than one which does not. The degree to which such relations should be incorporated, and the distances up and down inclusion chains or within coordinate arrays which should be searched, are problems which, like the discrimination of relations between terms in compounds, can only be settled by statistical investigation.

OPTIMUM LEVEL OF DISCRIMINATION

The analyses discussed above provide a set of terms (descriptors) which are linked in an information lattice by subordinate and coordinate relations, and linked in compound subjects by interlocking relations. The deeper the analysis is taken, the more discrimination between subjects can be built into a system.

As has already been stressed, there is no natural end to this process of analysis. The attributes of natural phenomena are of endless variety and uncountable number, and there are always more which can be drawn upon to discriminate more finely. The only criterion we can adopt to establish optimum levels of discrimination is the practical one of helpfulness in information retrieval. Two criteria are available.

The first is known as *literary warrant* and it is this: that if a given subject has appeared in the literature, and if it is desired to retrieve documents relevant to that subject, then it must be possible to represent the subject by the descriptors used in the system. This criterion affects in the first place the choice of descriptors (indexing terms). If a newly appearing subject cannot be represented by existing terms (or combination of terms), greater discrimination, in the sense of more terms, must be introduced into the retrieval system. Apart from this, literary warrant also supplies the data for establishing inclusion chains, coordinations in array, and interlocking relations for building into the information lattice.

The second criterion, and the controlling one in deciding on the optimum level of discrimination, we may call *user relevance*. There may be literary warrant for discriminating between the two compounds “Destruction of bacteria by dyestuffs” and “Destruction of dyestuffs by bacteria,” but in fact a searcher asking for one may find the other relevant, as each is an instance of the more general subject, “Destructive relations between bacteria and dyestuffs.” Again, the same searcher may find that the still more general subjects “Relations between bacteria and dyestuffs,” “Destruction of bacteria,” or even “Bacteria” all retrieve relevant documents. Exactly the same criterion of user relevance applies to the successively more general terms up an inclusion chain, or the successively more alien terms in a coordinate array. The relevance of discrimination between subjects in the use of a retrieval system must be decided by a study of that use.

Every retrieval system introduces some prior discrimination—if it did not, there would be no system. Descriptors are chosen, and lattice relations between them may be established. This prior discrimination is not a priori—it is based on literary warrant. It is quite justifiably assumed that discriminations which have been relevant to authors in the past will be, to a greater or lesser extent, relevant to readers in the future. The problem is how best to combine literary warrant with sensitivity to current user relevance and, in particular, how to build this sensitivity into the retrieval system, so that the system can "learn" the optimum levels of discrimination.

Two extreme solutions are possible and are currently practised. The first is to construct an information lattice, based on literary warrant, with a minimum of discrimination, so that the relevance limits are wide. Such lattices are used in some machine retrieval systems. As the collection of documents grows, so will the noise factor—the percentage of irrelevant material retrieved in a search. When the noise factor becomes too high, more discrimination is built into the system.

The alternative solution is to establish as detailed an information lattice as possible, building in all the terms and their relations encountered in a close study of the literature. This is the aim set in detailed faceted classification schemes. In card catalogue applications of such schemes, the *relevance limits* are set by the user: in the course of an actual search, he decides how far to pursue the inclusion and coordinate links in the lattice. For example, after examining the references (or the documents) retrieved at each step up or down an inclusion chain, he can decide whether further steps are relevant to his request, i.e., he can set his own optimum relevance limits. In man-machine retrieval systems, exactly the same procedure can be followed.

In machine systems which do not operate in such close contact with the ultimate user, it may be possible to "quantify" optimum relevance limits. Suppose that on average a system operated at 10 levels of discrimination (e.g., 10 steps per inclusion chain), and the machine is programmed so that, requested to search for subject S at level L it will retrieve all items marked S and those at levels L+1, L+2, L+3, L−1, and L−2 related to S. Given a sufficiently flexible machine, tallies for items rejected by the user as irrelevant could be fed back into the machine, which could record the rejected levels. Statistical examination of such records might lead the operator to re-programme the machine to omit, say, levels L+3 and L−2 in future searches. The machine might even re-programme itself.

Classification schemes and other schedules for retrieval systems are often closely based on literary warrant and techniques of subject analysis for this purpose are being worked out. But more thought is needed on how to make systems sensitive to newly emerging literary warrant, and adjusted to current user relevance. There is considerable scope for research in this field.

BIBLIOGRAPHY

D.D.ANDREWS and S.M.NEWMAN. *Storage and retrieval of contents of technical literature, nonchemical information,* U.S. Patent Office, Research and Development Report, 15 May, 1956.

J.E.L.FARRADANE. A scientific theory of classification and indexing, *J. Document.,* 1950, *6,* 83–99; 1952, *8,* 73–92.

W.P.K.FINDLAY, *Trans. Brit. Mycol. Soc.*, 1949, *31,* 106.

H.P.LUHN, *A statistical approach to mechanized literature searching,* International Business Machines Corporation, New York, Research Paper RC-3, 30 January, 1957.

J.W.PERRY, A.KENT, and M.M.BERRY. *Machine Literature Searching,* Interscience, New York, 1956.

S.R.RANGANATHAN. *Prolegomena to Library Classification,* 2nd edition. Library Association, London, 1957.

B.C.VICKERY. Systematic subject indexing, *J. Document.,* 1953, *9,* 48–57; The significance of John Wilkins in the history of bibliographical classification, *Libri,* 1953, *2,* 326–43; Common facets and phase relations, *Ann. Libr. Sci.,* 1957, *4,* 8–12.

Proceedings of the International Study Conference on Classification for Information Retrieval, ASLIB, London, 1957.

The Construction of a Faceted Classification for a Special Subject

D.J.FOSKETT

The Classification Research Group (C.R.G.) in London has been discussing for some years the theory of documentary classification, and several papers have been published which reflect the course of the discussions (1–8). Beginning with an explicit disavowal of allegiance to any one published system, the Group has considered the well-known schemes, both general and special, and the work being published by those in other countries who have also been studying the subject theoretically. It has not, unfortunately, had the opportunity so far of seeing the system developed in the U.S.S.R. on the basis of the philosophy of dialectical materialism.

While the Group has not been particularly satisfied with the development of the Colon Classification itself, we have nevertheless come to the conclusion that the method of facet analysis, first used systematically by S.R.Ranganathan, though sometimes occurring previously as it were by intuition, should form the basis of all forms of information retrieval. Vickery (5) has shown how several workers who ostensibly reject any form of classification, preferring mechanical sorting or other non-systematic coding, have actually begun to introduce, however reluctantly, some of the groupings that have long been commonplace features of classification schemes. In some cases, their reluctance is made evident by attempts to disguise these commonplace notions in weird and sometimes self-invented pseudoscientific jargon, supported, albeit unnecessarily, by masses of impressive mathematical diagrams and calculations.

The characteristic attitude of opponents of classification, particularly in the U.S.A., is that classification has been tried for many years and is demonstrably inefficient. This attitude cannot be justified, because "classification" is almost invariably taken to mean no more than one or two particularly widely used systems, namely, the Dewey Decimal Classification or the Universal Decimal Classification, and the Library of Congress Classification. There has been little or no attempt to make more than a superficial examination of these systems,

D.J.FOSKETT University of London Institute of Education, London.

and no examination of the basic theory of classification and systematic arrangement, despite the fact that Ranganathan himself spent several months in the U.S.A. a few years ago. For all the signs in subsequent literature, this visit might never have occurred. The issue is further confused by the fact that several proprietary sorting systems are now being offered for sale, and there is thus a number of conflicting interests to prevent objective research on a co-operative international basis.

Several systems have been constructed for special subjects by members of the Classification Research Group, and there has been substantial agreement on the method of making the schedules. The objects of this paper are (1) to give an account of how some of these systems have been made, so that the method can be tested by others, and (2) to point out some of the problems for which only a makeshift solution has so far been found. The paper does not necessarily represent the views of other members of the C.R.G., though it is based on its work.

BASIC FACETS

Two things are required for the construction of a classification system for a special subject: a knowledge of classification technique and an understanding of the subject itself. It has been shown again and again that both of these two qualities can be found in the same person, who need not be both a practising classifier and a practitioner in the subject—indeed this is such a rare combination that it cannot be expected. It is desirable, however, that expertise in both the method and the subject should be available, and if the system is being devised by one person, he should be able to consult where necessary those whose knowledge complements his own.

On the whole, my impression is that it is better for the system to be made by an expert in the method, able to seek help and advice from the expert in the subject. This is because the subject expert rarely pauses to consider consciously the systematic organisation of his subject (though this would be a useful discipline for him), whereas the classification expert knows the sort of pattern he should expect to find in a subject, and for most of the details of his analysis he can use the standard textbooks and reference works on it. Much of the “Sources of Hazards” facet in my classification for Occupational Safety and Health, for example, is based on the Model Code for Safety Regulations issued by the International Labour Office (8).

It is important that experts in both fields should consult at the beginning of the operation, since the first thing to be done is to discover the basic facets of the subject, and this requires some familiarity with the literature. The theory of facet analysis has been often worked over by Ranganathan and members of his Indian school, and he originally postulated five fundamental categories (Personality, Matter, Energy, Space, Time), which have since been augmented by the concepts of Rounds and Levels. These allow the first three fundamental categories to appear more than once, and enable the Colon Classification to cope with the many-faceted subjects characteristic of modern research.

It does not, however, seem to be necessary to keep within the limits of these fundamental categories in order to use the technique of facet analysis, and it is not always helpful to do so. Ranganathan has never given an adequate exposition of the basis of his categories; he adopted them more or less intuitively, though this has not prevented their being strikingly successful in the analysis of some subjects. But many people certainly do not find them easy to comprehend and use, possibly because they make a considerable departure from traditional classification theory. More recently, Vickery (9) has given a more systematic account of the derivation of categories based on explicit analysis of subjects, and it seems simpler to follow this principle rather than to try to relate every facet discovered in a subject to the abstract categories of Ranganathan. Of the three systems used to illustrate this paper, only one, Food Technology, is based on Ranganathan's categories, because it is intended to be used as a part of the Colon Classification. The facets in the other two subjects, Occupational Safety and Health, and Container Manufacture, are based solely on the analysis of the subjects; even so, the facets in Food Technology would certainly be the same if derived by the second method.

The analysis of a subject into its facets is in practice a fairly simple operation. The first step is to examine a wide range of the literature and to enumerate the subject of each article in a manner such as Farradane uses in constructing his analets (10,11). It soon becomes clear that certain Substances, Products, Parts, Reactions, Operations, Tools, Agents, and similar categories constantly recur, and they should be grouped according to their importance in the subject. In some subjects, particularly in Technology, relating the facets to Ranganathan's fundamental categories can be helpful; in others, particularly the Social Sciences, it may confuse rather than clarify the situation, which leads to doubts about the universal applicability of the procedure. Consider these subjects:

The lighting of underground roadways in coal mines

The protection of workers against ionising radiations

The examination and testing of dust masks

Some dermatologic aspects of the chromate problem

The guarding of machines used by blind workers

Registration of accidents in nuclear reactors

Examination of these and similar subjects that occur in the abstracts in the journal *Occupational Safety and Health* show that in this subject the following facets occur:

Special classes of workers (blind, etc.)

Industries: where hazards exist (coal mines, reactors)

Sources of hazards: the things causing danger (dust, chromate, radiation, machines)

Accidents and diseases: the results of the hazards (accidents, dermatitis)

Prevention: the means by which the worker is protected (masks)

Organisation and administration: human problems and methods of solution, arising out of the practice of Occupational Safety and Health; a miscellaneous facet whose existence is required by the literature of the subject (registration)

When the facets have been established, the next step is to enumerate their contents—the individual items in them—as far as possible. The advice of the subject expert is particularly useful here, but it is ultimately a more or less mechanical task, and most of it can be done by using standard reference works, whose contents lists, section headings, bibliographies, and indexes all contribute. It is important to remember, moreover, that this task can probably never be completed, and it is a waste of time to try to find out every single item for every facet; what is important is that the notation attached to the system should be able to code, in the appropriate place, the new items as they arise in the literature.

PROBLEMS OF SEQUENCE

It is at this stage that the question of sequence arises, both for the items within a facet and for the facets themselves.

SEQUENCE WITHIN A FACET

While the object of a classification system is to arrange literature in the most helpful sequence, it is well known that, within any facet, one sequence of items may be as helpful as another. Certain items may have outstanding importance, such as Mining in the Industry facet of Occupational Safety and Health, and these are probably most helpfully placed at the front of the sequence. Similarly, it has always been considered more helpful to place general headings before their subdivisions: Hoisting Tackle should precede Cranes, Crabs, and Winches; Dangerous Radiations should precede Infra-red and Ultra-violet, Alpha, Beta, and Gamma Radiations. Where there exists some criterion such as these, by which items can be arranged in one certain sequence, it is reasonable to use it. But it often happens that, while all the items in a facet bear the same relation to the main subject, they do not bear any relation to one another that would enable us to say that one item should precede or follow another. In the Sources of Hazard facet we have such items as Fire, Industrial Equipment, Electricity, Dangerous Radiations, Environmental Conditions. It seems reasonable to keep Electricity and Dangerous Radiations together, though there is no particular reason why one should precede the other, unless date of discovery may be admitted. But apart from these two, the sequence of the group has no particular significance: one is just as important as the other in its relation to the main subject.

Similarly, in Food Technology, we find such subjects as:

Hot air sterilization of food packs

Aseptic canning of milk

Gas packing of peanuts

Extraction of juice from Seville oranges

These give us the following facets:

Product (foodstuffs, milk, peanuts)

Part (juice)

Raw material (Seville oranges)

Operation (sterilization, canning, gas packing, extraction)

The items in the Product facet include Dairy Products, Sugar and Sugar Products, Cereals, Bakery Products, Edible Oils and Fats, Fruit and Vegetables, Meat, Fish. Apart from Cereals and Bakery Products, which are related and may be kept together, there is no significant criterion by which one sequence might be judged to be superior to another for these groups of products. The same is true for the subdivision Fermented Milks, of the division Dairy Products: Yoghurt, Kefir, Kumiss, Leben, Mazum, Gioddu, Dahi, Acidophilus Milk.

SEQUENCE OF FACETS

The sequence laid down by Ranganathan (P,M,E,S,T) for the citation of facets in a class number was originally based on his observation of how readers actually preferred their literature to be arranged. In Class E, Chemistry, and Class 2, Library Science, the sequence first chosen had to be changed because it turned out to be unhelpful, and this observation led to an analysis of the reason behind this intuitively chosen sequence (12). The principle eventually enunciated was named the Principle of Decreasing Concreteness, and the “proof” assumes that a Matter facet is less concrete than Personality, Energy less Concrete than Matter (13). Ranganathan chose reasonably convincing examples to illus trate this Principle, and it may be acceptable if we limit our discussions to the terms of his own categories. In developing schedules for some subjects not in the Colon classification, such as Food Technology, this sequence is logical and helpful; it results in the literature being gathered together at the various groups of foodstuffs, and this is in fact the way the food industry usually works, rather than according to processes. It appears at first sight that there are exceptions, such as "canners," who are food processors practising food sterilization by heat, and whose primary interest would therefore seem to be in the process. But a closer look at the industry shows that actually there is a considerable amount of specialization, according to the food groups. There are fruit and vegetable canners, meat canners, fish canners; these are the main groups, and only the largest firms go in for more than one group. Their factories are located to be near the growing area, their processing equipment is designed for the product, even their cans are different for the different foods. So that even in cases like these, the *most* helpful sequence for the documents turns out to be the one in which the product precedes the process.

But when we turn to a subject such as Occupational Safety and Health, it is not so easy to relate Industry, Hazard, Diseases, Preventive Measures, to the P,M,E,S,T formula. In what sense can it be said, for example, that Tenosynovitis and Frostbite are *less concrete* than Bessemer Converters and Static Electricity? It is possible (indeed it has been shown to be probable) that those who are steeped in the lore and practice of P,M,E,S,T can sense this relationship of more to less concrete even in the most obscure sets of facets. But so far there has not appeared an adequate explanation of how to specify the quality of concreteness, and this is obviously a point that needs further investigation.

Other suggestions have been made; W.G.Stiles has proposed decreasing specificity or increasing probability, implying that the more specific a subject, the less frequently it occurs. This may often be true—it probably is for books —but at the documentation level there must be far more articles dealing with specific than with general subjects. Indeed, it is this very characteristic of periodical articles that has been one of the most important factors contributing to the present crisis in classification. This is another point that needs more detailed study.

J.Mills has a more pragmatical criterion: that of purpose or use. If the items in one facet comprise the end-product or result of the items in another, the former should precede the latter. In Food Technology, Evaporated Milk is the product of the Evaporation of the raw material Milk, and the sequence is Evaporated Milk—Milk—Evaporation, or Product—Raw Material—Process. As we have seen, this corresponds exactly to Ranganathan's Personality—Matter—Energy.

The evidence indicates that we must be fairly close to a sound explanation of facet sequence, since those so far given justify more or less the same sequence. But this leads to the suspicion that we have not yet discovered the actual basis for determining the sequence; what we have done is to find a justification— reasonable enough—for a sequence that has been chosen by instinct. Why, then, do we all seem to arrive at the same, or nearly the same, result? Is it simply because all our instincts are sound? Even if it is, we should be able to find a scientific basis for it.

The reason why this point has to be decided, of course, is that the sequence of facets is what determines the manner in which the documents are grouped. Suppose we have a subject with a series of facets A,B,C,D,E: then articles whose subjects are represented by AB, ABC, ABCD, ABCDE, AC, ACD, and so on, will all be collected at A. This means that the information on B,C,D,E, in these articles will be found at A, and the reader interested in them will have to look in more than one place in order to find all he wants, since BC and BCD will be shelved at B, and so on. The phenomenon of distributed facets has often been discussed, and these approaches have to be taken care of by an alphabetical index constructed by chain procedure. Consider the following example:

GUARDS	Eg
GUARDS:HAND INJURIES:PRESSES	CgfDbbEg
GUARDS:HOISTING TACKLE	CbnEg
GUARDS:PRESSES	CgfEg
GUARDS:WINCHES	CkjEg

Eg is the number for GUARDS, but information on Guards is scattered according to the type of machine guarded; the Source of Hazard facet precedes the Prevention facet. If the order of facets were different, information on Guards might be grouped, but then information on Winches and Presses would necessarily be scattered.

The actual criterion that determines the order in which facets shall be cited is this: supposing we have a subject with *x* facets, and a series of articles in which all are mentioned, at which one do we wish the literature to be grouped? If facet A is the most important, or the one most commonly sought, then the reader will best be served if all articles containing an item from facet A are filed together. Thus in Occupational Safety and Health, I have made Industry the first facet; but if an article deals with a special class of worker, such as Old Workers or Amputees, the whole content of the article is influenced by this fact, and this will be true for all articles (or the vast majority) in which these special classes of workers are mentioned. This leads to the conclusion that this category must actually take precedence as the dominant factor, even more important than Industry. The fact is, of course, that it is the worker who is the most important factor in all of this subject; the health and safety of the worker, not of the industry, are being considered. But it would be a waste of index and notation to specify "normal" workers in every case, and this is why it is possible to include Industry as well as Special Classes of Workers in the first facet.

This criterion can be applied to each facet up to the last two; once the more important of these is determined, there is no more to be said of the other. At each step, the preceding facets are omitted (because they take priority), and the question of grouping the literature answered.

This matter of "literary warrant," then, must be the basis for deciding on facet order, and the conscious or intuitive recognition of it has led to the great similarity in facet orders arrived at by other writers. It is necessary to be perfectly clear on the point, however; mere explicit enunciation of the criterion does not help us to find a principle to guide us in all subjects, and this is what we really want.

ALPHABETICAL SEQUENCE

Since the classic paper of Pollard and Bradford (14) destroyed for ever the myth of the superiority of an alphabetical sequence of subjects, there has been a tendency to exaggerate in the other direction and to reject alphabetical sequence at all costs and in every case. As I have already remarked, however, within a facet there is often no particular reason why one sequence should be preferred to another, and alphabetical arrangement has at least the merit of requiring no notation in a classified sequence. Generally speaking, I should not expect alphabetical sequence to be satisfactory except where the items to be arranged do not bear any particular relation to each other.

The main drawback to the use of the alphabet is often said to be that its ease of use is restricted to one language, but clearly each country can arrange the documents according to its own language where no sequence-indicating notation is involved that is an integral part of the classification system itself. This would perhaps not apply to an internationally used bulletin of abstracts, but since these are accepted in their language of origin anyway, this does not seem to me to be a serious drawback. Wherever there is no preferred sequence, then, alphabetical arrangement can be adopted without qualms, at least until a better arrangement suggests itself in practice. No notation need be allotted, because it would merely duplicate the already existing code symbol, the alphabet.

SOME PROBLEMS OF SPECIAL CLASSIFICATIONS

PRIMORDIAL SCHEDULES

One of the most difficult problems arising out of constructing a special classification system is the impossibility of "isolating" the subject from related fields of knowledge; at almost every step instances occur in which part of another subject has to be "borrowed" in order to complete the special system. This has led to proposals to draw up "primordial schedules," for example of Raw Materials, and Manufactured Commodities, which could be drawn on for use in any special system. The obvious attraction of this idea is enhanced by the possibility of using a distinctive notation which would acquire considerable mnemonic value. It would call for a special effort, but there exist already certain lists, for Customs purposes for instance, that might be used as starting points and save at least some of the work.

DEPENDENT FACETS

Another problem arises when the items in a first or second facet, such as Products or Raw Materials, bear no significant relation to each other, because it may mean that the items in subsequent facets, such as Processes, share this lack of relation. In Food Technology, for example, the processing operations used in the cereal products industry are, in the main, different from those in the fruit and vegetable or the meat industry. In other words, while the actual category (Operation) is common to all the products, the items in it vary with individual items from the Product facet. Two solutions appear to be possible: either to make an extended Operation facet listing all the processes for all products in the one sequence, or to make a separate Operation facet for each product.

The advantages of the first course are that there are usually some common processes, which need to be itemised only once; there are some operations which, though different for each product, can be described in a sufficiently abstract way to fit several products, for example "removal of non-edible portions" (such as husk, chaff, peel, core, bone); it is always possible that an operation used for one product but not another may in due course, with the advance of technology, be applied to the second, when it presents no problem whatever because both parts of the new subject are already present in the classification.

A good example of this can be given from schedules for Container Manufacture, in which subjects like these occur:

Gaskets for self-heating cans

Decoration of tinplate fish cans

Soldering of paint can ears

Cleaning of aluminium extruded tubes

From these, and others like them, we deduce the following facets:

Product (cans, fish cans, paint cans, extruded tubes)

Part (ears)

Material (tinplate, aluminium)

Operation (soldering, cleaning)

From the beginning, the Material facet included the items Metal and Paper, and a single Operation facet was used, which included the item Drawing, because drawn metal containers were often written about. At that time, the drawing operation had not been applied to paper or cardboard; but only a few years after the scheme had been drawn up, drawn paperboard containers were actually produced and became the subject of reports. The scheme was faced with no need for addition, because both items were already in it, and could be linked together, because of the faceted structure, with no more trouble than an already well-known subject.

The advantages of separate secondary facets are that they may result in a shorter notation; they are easier to use because each item in the facet is relevant to the product being considered and is readily recognised by the classifier because it can be given its distinctive and unambiguous name, and does not have to be described in a more abstract manner suitable to several products.

Separate facets may prove to be necessary for certain categories which appear at first sight to be common to many subjects. Properties and Faults are obvious examples, where so many substances are involved that a truly common facet would be of very great length, involving a long notation. Control is an example of a category which was at first thought to be a common facet, but in which the subdivisions actually vary considerably according to the items specified from the previous facets.

PRODUCTS AND OPERATIONS

The formula Product—Raw Material—Process—Agent for facet sequence appears to be satisfactory for several technologies, but needs further detailed testing, and the distinctions between Product and Raw Material have to be kept clear; this is not so straightforward as might be imagined, and it leads to important questions that must be answered.

Consider first the subject Food Technology. There is an item in the Product facet, Dairy Products, of which the first subdivision is Milk and Special Milks.

Milk thus stands as an item in the Product facet. In the Operation facet we have the item Evaporation, and the Evaporation of Milk produces Evaporated Milk, for which Milk is the Raw Material. Milk must thus stand as an item in both Product and Raw Material facets. But we can go further than this: also in the Operation facet we have the process Canning, and Canned Evaporated Milk is another product, for which Evaporated Milk is the raw material. How far should this analysis be taken? Should there be a place, with notation, in the Product facet for every processed foodstuff in its final form? To leave Canned Evaporated Milk to be represented by Evaporated Milk: Canning means that it is notationally indistinguishable from The Canning of Evaporated Milk, and it may be that this does not matter, since the actual literature may be better grouped at one number rather than at two probably consecutive numbers.

There are two objections to this. Firstly, we may have to take the matter yet another stage further, to deal with literature on Spoilage of Canned Evaporated Milk, which is quite a different subject from Faults in Canning Evaporated Milk. Both these subjects would receive the same class number. Secondly, there are some foodstuffs that may be processed by any of several methods, each producing a different end-product: canned peas, dried peas, frozen peas, canned frozen peas, canned dried peas. Should these be separated into individual items in the Product facet, and if so, how, and in what sequence?

Similar examples can be quoted from Container Manufacture, where the formula Product—Material—Operation—Agent also applies. The Material facet includes, as we have seen, Tinplate, Aluminium and Lacquers, the Operation facet includes Coating, and Lacquering. Lacquering of Tinplate and Aluminium is not equivalent to Lacquered Tinplate and Lacquered Aluminium, and it would therefore seem that, as containers are made from these materials, they must also appear in the Material facet. We must further distinguish between Hot-dipped Tinplate and Electrolytic Tinplate, and so make yet further subdivisions for Lacquered Hot-dipped Tinplate and Lacquered Electrolytic Tinplate.

Lacquers for Tinplate raises a different problem. It is clear that this is a different subject from either Lacquering of Tinplate or Lacquered Tinplate, and must be distinguished from them. But the problem here is one of priority: should documents on Lacquers for Tinplate be collected at Lacquers or at Tinplate? The problem is even further complicated by the history of the subject. With the introduction of Electrolytic Tinplate, interest was concentrated on the new material, so that Lacquers for Electrolytic Tinplate seemed to be best collected at Electrolytic Tinplate; but now that it is in common use, interest has reverted to the Lacquers, and the literature would be better collected there.

This would actually seem to be a case of what Ranganathan calls Bias Phase, in which the documents should be classified at the subject which is affected (Lacquers) rather than at the one which affects (Tinplate).

General and special classifications

Although there are some writers in this field (mostly the vendors of proprietary systems) who maintain that only a specially prepared system will serve for a special subject, yet it is impressed on anyone who makes a special classification that the advantages of having a general system would be enormous. The main drawback at present is that there is no general system that is at all satisfactory for documentation, and special systems must continue to be made, despite the appalling waste of time and labour, until some organisation will assume the responsibility for making a new general system. This conclusion has been reached after long discussion in the Classification Research Group, which considers that such a system could now be made by the method of facet analysis. The system, by its very existence, might well solve some of the problems I have described; for example, the study of lacquers does not properly belong to the subject of Container Manufacture, though it has to appear in a special system because the literature demands it. In a general system all documents dealing with Coating Materials (Lacquers, Varnishes, Inks, Waxes, etc.) as the primary focus of attention would be collected into a class for that kind of material.

This is one aspect of the principal advantage to be gained from a general system, which is that one does not have to list for each subject all its possible ramifications, that is, all the other subjects which may at some time be related to it, or even form an integral part of it. Phase relations such as Bias and Tool are obvious instances. Books on mathematics are written with a bias towards a number of professions, engineers, accountants and so on, but one would not on that account wish to list all conceivable variations of this kind in a classification for books on mathematics. Where the related subject may form an integral part of the special subject, however, it may be desirable to specify at least some of its items in a special facet. A good example of this is the Agent facet of some subjects, such as Spoilage Organisms in Food Technology. One of a great many organisms may be responsible for food spoilage, and it is therefore desirable to be able to draw on a general schedule of micro-organisms such as would form the primary facet in the subject Microbiology. This is a clear case in which it would be most uneconomical to make a second list for Food Technology; and of course Food Technology is by no means the only subject in which micro-organisms play the role of agents in some biological process. But a very few organisms are responsible for most of the spoilage in each food group, and it may be desirable to list these in an Agent facet, either a general facet for the whole subject, or separately for each food group. This is what Ranganathan calls the Favoured Category Device, although these are not categories but individuals items. Thus *Clostridium botulinum, Byssochlamys fulva, Salmonella* types, and so on, would be given brief numbers and would appear at the front of the schedule; and an instruction added, "For other organisms use the number from Class X Microbiology."

This is a compromise solution which relies on the background of a general system for the rarer items, but includes the "favoured" items in a special facet added to the main facets. It will also meet the criticism that use of a general system often means that the items selected are not in the preferred order for the special subject, that the most important items, which would preferably be at the front, may be in the middle of a group of relatively minor items that occur only rarely in the literature of the special subject. In other words, the preferred order for micro-organisms in the class Microbiology is not the same as the preferred order in the class Food Technology. But this is in practice only important for frequently occurring items, which can be "favoured" by being listed in an Agent facet in the special subject. The sequence of items that occur rarely is of no particular consequence, and it may well be useful if these are in the same sequence as in their parent class, because this will be the most helpful sequence for specialists in that class, and so express relationships that help to characterise the items listed.

CHAIN INDEXING

The technique of chain indexing has often been described, and it is not necessary to do so again here. It must be mentioned, however, since an index constructed by chain procedure forms an essential part of a faceted classification, and opponents of classification often assume either that no index would be part of such a system, or that its compilation would be a task requiring much time and labour. It cannot too often be stressed that neither of these assumptions is true, and that they must no longer be allowed to obscure a perfectly straight-forward situation.

The principal reason for introducing chain indexing here, however, is that the technique does exert an influence on the construction of schedules for a faceted system. Briefly, the technique is to index each facet of the specific subject in turn, beginning with the right-hand facet and proceeding from right to left. Thus the index entries for the subject "Lighting of underground roadways in coal mines" would be:

LIGHTING: ROADWAYS: MINES, COAL

ROADWAYS: MINES, COAL

MINES, COAL

COAL MINES

The object of the technique is to effect a symbiosis between classifying and indexing, and it means that the choice of subject headings no longer depends on the flair of the indexer or on some arbitrarily chosen list of subject headings, crutches that seem to be particularly popular in the U.S.A., if one may judge by the number that are published there. But if the technique is to be effective, it requires that every facet of the specific subject of a report or periodical article should be expressable in the classification system and distinguishable by means of the notation. This is the reason why it is so important that the system must be based on literary warrant. If it is not, there will be facets in the literature that cannot be expressed by the system, and they will consequently not appear in the alphabetical index. It is as a cure for precisely this fault that chain indexing has found so much favour in countries where the technique is known.

NOTATION

Some of the thorniest problems in classification are basically problems in notation, which indicates its importance in spite of the fact that it is no more than a device for mechanising the sequence of items in schedules and documents on shelves. Most of the recent discussions on coding have been directed towards machine sorting rather than classifying, but Ranganathan has been systematically developing his “artificial language of ordinal numbers” for many years, and some entirely new ideas have recently been suggested to the Classification Research Group by E.J.Coates (15). My systems for Occupational Safety and Health, and Container Manufacture, have used some of Coates' ideas, but I have numbered each facet separately, and have not incorporated all but the last facet into a single “retroactive” alphabet. The best example of Coates' notation is that used for the classification of the *British Catalogue of Music*.

There are two qualities of the first importance in a notation, and unfortunately they are to some extent in conflict: brevity, and ease of recognition. I am inclined to think that since the primary function of a notation is to mechanise sequences, its primary quality should be to make sequences readily recognisable, even at the expense of brevity, which is often considered to be the most important quality. The main reason is that the briefest possible notation, such as that now being sought by Ranganathan, is obliged to use different species of symbols in the same sequence, for example within a single facet. This means that the sequence of the symbols has to be arbitrarily laid down, since there is no traditional and recognised single sequence that includes more than one species of symbol, say, letters and numbers. On the other hand, brevity is highly desirable, and the advantage of roman letters over arabic numerals in this respect seems so marked that it more than compensates for the greater ease of recognition of numerals. Thus for numbering items in each facet I have used lower case roman letters. Incidentally, now that China and Japan are proposing to adopt a roman script, one of the chief drawbacks to this species of symbol will lose some of its force.

Different species of symbol can play a part, however, in indicating a different type of item, and here they are actually helpful in showing the change in the nature of the item symbolised. There are several parts of a notation that have to be distinguished from each other, and several species of symbol available; the problem is to fit them together in a manner that achieves a readily recognisable sequence without too much loss of brevity.

The parts are: facets, items in facets, common subdivisions, phase relations. It might also be useful to have a distinctive symbol for the items in Space and Time facets, as these are likely to be wanted at any point, and should therefore be easily addable to any number.

There are several species of symbol available, but only three of them have a readily recognisable sequence: arabic numerals, upper and lower case roman letters. I do not consider that italics can be admitted, as they run too great a risk of confusion with ordinary roman letters, and are not easily memorised as distinct from them. Other symbols are those used in the Universal Decimal Classification and the Colon Classification, but a sequence for them has to be arbitrarily chosen, and this definitely arouses opposition, which Ranganathan is inclined to underestimate.

The notation chosen for Occupational Safety and Health, and Container Manufacture, uses upper case letters as facet indicators, lower case for terms in facets, a dot to introduce common subdivisions, and parentheses for phase relations. Space facet needs no indicator where, as in Occupational Safety and Health, an already existing schedule symbolised by arabic numerals can be used. Time could well follow U.D.C. practice and insert the date in inverted commas; this is not often required, and the lack of brevity is compensated for by the immediately recognisable, meaning exactly what it says.

This produces symbols like these:

The lighting of underground roadways in coal mines	Bec Cb Epd
Accident research in the iron and steel industry	Bx Dbb.f
Treatment of workers suffering from coal-workers pneumoconiosis	Ker Mg(Bec)

With this notation the facets are clearly separated from one another, each facet is distinctively marked, the terms in it follow a simply numbered sequence, common subdivisions and phase relations are readily recognisable as being outside the normal sequence of facets. The notation is thus quite uncomplicated, and I have tried to use the non-sequential symbols in a manner that does not strike the user as strange. This has no doubt meant some loss of brevity, but once again is more acceptable. Many discussions about the Colon Classification leave me with the impression that the complexity of its notation is one of the greatest barriers to its more widespread use, and when a new general classification scheme is made (as it must be), its makers would do well to give consideration to this quality of familiarity and consequent acceptability in its notation.

Since Melvil Dewey introduced the use of a decimal notation, it has often been assumed that notation should express the hierarchy of the terms of the classification, though Brown and Bliss both depart from the practice. Certainly some users at least expect notation to reflect hierarchy; but there is no great difficulty in comprehending a non-hierarchical notation, and it has the considerable advantage of making the notation for subdivisions shorter by one or two symbols for many terms, and by more for the more specific subdivisions.

One of the major drawbacks to non-hierarchical notation might be difficulty of application to punched cards (for which, incidentally, a faceted classification is particularly well suited), because if an alphabet is spread over a main division and its subdivisions, the symbol for the subdivisions would not include that for the division itself. Thus if Cc stands for Fires and Explosions, Ccb for Fires, and Cf for Explosions, sorting by the symbol Cf would not bring out the more general references coded Cc, in the way that sorting for 123 brings out 12 as well. But since the inclusion of 12 in 123 is after all no more than a mathematical convention, there seems to be no reason why a similar convention should not be established for punched cards, by which the symbol Bb contained as its subdivisions the symbols Bc to Bg, and so on. It would not even be necessary to keep the same convention for every class and division, but obviously the application of such a radical departure from established practice would require much care.

CONCLUSIONS

This paper may seem elementary, in parts if not throughout, to those who feel that no paper on information retrieval can be significant if it is totally lacking in formulae, equations, and even diagrams. On the other hand, much of the writing in this field shows a complete ignorance of the technique of facet anal ysis, and there has been a strongly expressed wish from many quarters for a plain and straightforward exposition of the method of schedule construction for a faceted classification. In the Appendix, I have given extracts from all three of the systems quoted as examples in this paper.

The discussions of the Classification Research Group, and the writings of its members, leave no doubt that classification and indexing are sciences of considerable complexity, and this is inevitable because their purpose is to bring some order into the enormously complicated mass of modern documentation. A complicated tool will, if efficiently constructed, do a much more skilful job than a stone axe, and we do not expect complicated tools to be made without detailed specifications. But recent publications of certain groups of theorists seem to be receding from the basic purpose, which is to arrange documents and indexes in a helpful sequence, and as a result many librarians and documentalists who have a real need for classification are driven away and take hopeless refuge in dictionary catalogues and lists of subject headings.

Yet all over the world, work is going on to form special systems or to try to keep older systems going by means of repairs, amendments, new editions; a vast machinery exists to prolong the life of the Decimal Classification and the U.D.C., which G.Cordonnier has with justice described as “un monstre pré-historique.” At the Brussels Congress of 1955, I urged that the trend towards faceted classification should be recognised, and further research sponsored; as a result, the International Study Conference on Classification for Information Retrieval was held at Dorking in May, 1957. It was attended by representatives from seven countries, and reached a remarkable measure of agreement on what are the basic requirements in this field. The Washington Conference should take the next logical step forward and establish the necessary machinery for creating a new general classification system by organised international effort.

REFERENCES

1. *Classification Research Group Bulletin,* 1956–.
2. *Library Association Record,* Vol. 55, 1953, pp. 187–188.
3. *Library Association Record,* Vol. 57, 1955, pp. 262–268.
4. VICKERY, B.C. Classification of chemistry. *Abgila,* Vol. 3, 1953, pp. 11–24.
5. VICKERY, B.C. Developments in subject indexing. *Journal of Documentation,* Vol. 11, 1955, pp. 1–12.
6. FOSKETT, D.J. Modern trends in classification. International Conference of Libraries and Documentation Centres, Brussels, 1955. Vol. IIb, pp. 96–102.
7. CLASSIFICATION RESEARCH GROUP. Bibliography of papers on classification and allied subjects. *Journal of Documentation,* Vol. 12, 1956, pp. 227–230.
8. *Proceedings of the International Study Conference on Classification for Information Retrieval,* Dorking, 1957. London, Aslib. 1958. A short account of the Con

ference, with the text of the "Conclusions and Recommendations," was published in the *Library Association Record,* Vol. 59, 1957, pp. 304–306.
9. VICKERY, B.C. *Classification and indexing in science.* (Forthcoming).
10. FARRADANE, J.E. *Journal of Documentation,* Vol. 6, 1950, pp. 83–99.
11. FARRADANE, J.E. *Journal of Documentation,* Vol. 8, 1952, pp. 73–92.
12. PURANIK, K.D., in *Depth Classification,* edited by S.R.RANGANATHAN. Delhi, Indian Library Association, 1951, p. 69.
13. RANGANATHAN, S.R. *Philosophy of library classification.* Copenhagen, Munksgaard. 1951. pp. 54–61.
14. POLLARD, A.F.C., and BRADFORD, S.C. The inadequacy of the alphabetical subject index. Aslib. *Report of the Proceedings of the Seventh Conference,* 1930, pp. 39–54.
15. COATES, E.J. *Classification Research Group Bulletin,* No. 1, 1956, pp. 13–19; No. 2, 1957, pp. D1−D19.

APPENDIX SAMPLE SCHEDULES

1. *Occupational Safety and Health*

Occupational Safety and Health: General	Facet A
Special Classes of Workers, Industries	Facet B
Sources of Hazards: Fire, Machinery, etc.	Facet C
Industrial Accidents and Diseases	Facet D
Preventive Measures, Protection	Facet E
Organisation, Administration	Facet F

FACET B

b	Special classes of workers
bb	Women and young workers
bh	Old and handicapped workers
bm	Amputees
bn	Blind
d	Agriculture, forestry, fisheries
e	Mining, quarrying
eb	Products
ec	Coal
em	Quarries
g	Oil and natural gas
h	Nuclear reactors
v	Armed Forces
x	Manufacturing industries

FACET C

b	Dangerous places in general
c	Fires, explosions
cb	Fires

f	Explosions
g	Industrial equipment and processes
ge	Machines
h	Hand tools, power tools
j	Furnaces, kilns, ovens
jb	Blast furnaces
jp	Brick and pottery kilns
p	Electricity
q	Dangerous radiations
x	Environmental conditions

FACET D

b	Deaths
bb	Accidents
be	Injuries
bd	Fatigue
bg	Diseases
c	Syndromes
e	Respiratory system
eb	Asthma
er	Pneumoconiosis
p	Nervous system
pk	Central nervous system

FACET E

b	Provision of preventive measures, protection
c	Alarm and detection systems
d	Escape means
g	Guards, fences
n	Personal protective equipment
nd	Clothing
nf	Heat resistant
t	Training and education
v	Medical supervision and surgeries

FACET F

b	Safety and Health Organisations
c	Health and social problems
g	Work problems
gf	Home work
gg	Night work
gm	Seasonal work
m	Legal problems
p	Notification and registration
q	Compensation

2. *Container Manufacture*

Container Manufacture: General	Facet A
Products	Facet B
Parts, Components	Facet C
Materials	Facet D
Operations	Facet F

FACET B

b	Metal containers
bb	Open Top cans
m	General Line cans
r	Non-metallic containers
s	Cartons
sz	Bottles
t	Flexible packages
v	Laminates

FACET C

b	Cylinders, bodies
c	Ends
h	Valves
q	Caps
qb	Screwcaps
v	Joints, seams

FACET D

b	Metals
c	Tinplate
a	Aluminium
g	Paper and board
k	Plastics
kc	Polythene
l	Film
q	Cork
qg	Glass
v	Coatings, decoration

FACET F

b	Analysis
c	Coating
d	Printing
f	Forming
fj	Extruding
fk	Impact extruding
fm	Moulding

g	Assembling
gc	Soldering
gt	Glueing
m	Testing, inspection
t	Coding
v	Storing

3. *Food Technology: an expansion of Colon Classification, Class F 53*

PRODUCT FACET

F531	Dairy products
2	Sugar
3	Cereals
4	Bakery products
5	Edible oils and fats
53	Individual oils, divided like J 5, e.g., F53571 Olive oil F53582 Coconut oil
59	Fats
6	Fruit and Vegetables
7	Fruit, divided like J37, e.g., F5372 Citrus fruits F53732 Melon
79	Nuts
8	Meat
91	Fish

PARTS FACET

F 53,1	Stalk
2	Skin, fur
4	Fibres
5	Juice

MATERIALS FACET

Can be taken from Class J Agriculture, and Class λ Animal Husbandry

OPERATIONS FACET

F 53:1	Preliminaries, preparation
12	Delivery
16	Cleaning, washing
2	Processing
21	Grinding, milling
23	Expressing
28	Evaporation, distillation
281	Evaporation

282	Distillation
3	Preserving
31	Fermenting
33	Heat treatment
332	Sterilizing, including canning
3323	Aseptic canning
4	Semi-preserving (divide like F53:3)
7	Packing, despatch
8	Storage
92	Testing, inspection
94	Spoilage
95	Hygiene, protection

On the Coding of Geometrical Shapes And Other Representations, with Reference to Archaeological Documents

J.C.GARDIN

1. INTRODUCTION

The object of the present paper is to describe a method of analyzing various representations, geometrical or otherwise, related to their storage and retrieval.

By “representation” we mean either abstract designs or figurative constructions, such as may be found in scientific or technical documentation. They may be two or three dimensional, and they may constitute either the very subject-matter of research, or only a graphic rendering of the various data under consideration.

2. METHOD

The method consists in breaking down the representation into various *elements,* and in expressing in different ways the *relations* according to which those elements are assembled.

The individuality of a representation is thus made explicit through a particular *combination* of terms, chosen from a limited set, to indicate both constituent elements and relations. Those terms are derived by making an analysis of the components of the representations being studied. They have no intrinsic value from the point of view of scientific understanding or, if so, only incidentally; their sole purpose is to provide a means whereby a certain type of document, namely graphical, can be analyzed and so be made accessible to search.

The criteria which are used for the selection of such terms are therefore purely operational. The objective is to devise the most *economical* system of

J.C.GARDIN Centre National de la Recherche Scientifique, Paris.

elements and relations through the various combinations of which specific representations can be defined with an acceptable degree of approximation, from the documentation point of view.

A compact code is thus constituted; it provides a way of expressing with a relatively small set of elementary, non-ambiguous features, a very large number of aggregate percepts, which have names too vague or no names at all in common usage.

The formulas which are obtained by analyzing representations in terms of that code may be considered as names, in so far as they can be made pronounceable through a reasoned attribution of phonemes (or morphemes) to each term in the code. The merits of such designations tend to be threefold: *objective,* i.e., conform to fixed standards of description, irrespective of personal appreciations; *international,* i.e., independent of national differences in the process of naming identical entities; *analytical,* i.e., capable of being broken down into several terms, which makes for a more compact storage, and a more flexible retrieval of information.

3. DEVELOPMENTS ON DIFFERENT LEVELS

The method which has been outlined is applicable on several different levels, according to the degree of complexity of the representations under study.

3.1. ABSTRACT DESIGN

Abstract design on a plane surface is the simplest case. The code is then essentially composed of terms borrowed from two-dimensional geometry, some designating, for instance, the most common figures (circle, triangle, etc.), while others refer to various ways of combining them in different patterns (intersecting, interlocking, radial, etc.).

A specific design is thus considered as the result of *n* successive "operations," carried out in a given order on one or several elementary "signs." Operations and signs are to be determined, not through deductive reasoning, as in the case of pure geometry, but through an empirical study of the particular kind of design at hand so as to produce the most economical system of analysis.

An example is given under "Abstract Ornament" (4.1).

3.2. THREE-DIMENSIONAL CONSTRUCTIONS

These can be analyzed in the same way, following, however, different procedures according to the kind of data under investigation.

3.21. In the simplest case one will have only to add a new term to the two-dimensional formula, so as to transform the corresponding plane design into a regular three-dimensional construct.

3.22. For more complex shapes it may prove convenient to choose three-dimensional figures as the fundamental units in the code, to be used in connection with a reduced set of operations.

3.23. Lastly, for constructions which cannot easily be reduced to such simple transformations, a good procedure consists in analyzing successively various cross-sections, either for the whole or for some predetermined parts of the object. An example of that method is given under "Artifacts" (4.2).

3.3. ICONOGRAPHICAL COMPOSITIONS

As representations grow more elaborate, and as they tend to develop into truly iconographical compositions, it becomes convenient to shift again the level of reference: the units of description are no longer geometrical elements —two or three dimensional— but figurative motives which are usually well understood under their common designations: a man, a building, a tree, etc.

3.31. Each motive may appear under different forms, which do not always receive specific names. The "building," for instance, is sometimes termed a hut or a skyscraper, but many times also a four-floor house, a flat roof house, etc., according to its shape, its height, etc. The analyst has therefore to consider the various manifestations of each motive; and his first task is to try to reduce the total set of variations, for all the motives taken into consideration, to a limited number of abstract terms, which take up different meanings according to the particular motive with which they are combined.

The vocabulary is thus made of two sections: a *lexicon,* where elementary motives are entered under their common name, and a *morphology,* where a list is given of the various complementary terms which serve to specify the form of each motive in the representation. (See the example in Section 4.31).

3.32. One has also to indicate the relations which prevail between the various constituents. Relations are of two kinds, logical and topographical.

By *logical relations* we mean the sort of interactions which bind different beings to one another in the same picture: *A* kills *B, A* sits in front of *C,* etc. Such relations are not made wholly explicit through verbs, nor through any other terms referring to an action, since the combination of three elements— two nouns *A* and *B,* and a verb ×—is usually ambiguous: A×B, *or* B×A (e.g., *A* kills B, or *B* kills *A*). In all cases where *A* and *B* can be substituted for each other with reference to ×, one has therefore to specify in addition to × the part respectively played by *A* and *B* in the corresponding action. Declensions provide in this respect a convenient solution: *A* (nominative, or subject)×*B* (accusative, or object), as opposed to *A* (object)×*B* (subject).

An example of a code with inflected terms is given under "Logical Relations" (4.32), for the analysis of iconographical documents in the field of archaeology.

The second kind of relations to be specified, *topographical,* concerns the relative position of the various beings which are associated in a given representation. Here again, the grammatical terms which usually convey that information (mainly prepositions) would not entirely satisfy the purpose, for they could often combine with any of the surrounding elements to provide entirely different meanings. To obviate the difficulty a few more grammatical "cases" have to be used, such as the locative and the instrumental, which, associated with definite nouns, play the same part as prepositions, without leaving any ambiguity as to the way in which the representation is actually organized.

Sometimes, for the more common disposition of subjects and objects, a single term may convey all the information which is needed as to the relative disposition of the various elements:

Term p $S \rightarrow O \leftarrow S$. Ex. Two heroes *(S)* striking a lion *(O)* standing between them.

Term q $O \leftarrow S \rightarrow O$. Ex. A hero *(S)* grasping to his left and to his right two lions *(O)*

Term r $S^2 \rightarrow S \rightarrow O$. Ex. A hero ($S^2$) striking a lion *(S)* which is attacking a lamb *(O)*.

S and *O* indicate respectively any subject or object in an iconographical sentence.

3.33. The more complex figurative representations can thus be analyzed in the same way as geometrical designs or functional objects. By combining terms pertaining to three different categories, lexicon, morphology, and grammar, the analyst is able to describe a very large number of iconographical themes, with a degree of approximation which he can set high or low, according to the range of his investigation. Yet the total number of terms, in the whole code, remains relatively low (cf.parag.43.3).

4. APPLICATIONS TO A SPECIFIC FIELD, ARCHAEOLOGY

The theoretical developments which form the subject of the preceding paragraphs have found an application in a systematic experiment carried out under the auspices of the French Centre National de la Recherche Scientifique for a specific field, archaeology.

The aim was to show that archaeological data, which are now scattered in countless publications and collections, could be made readily accessible to search through an appropriate coding of the information, independent of dif ferences in national languages or private idiosyncrasies, and through the recording of the coded information on punched cards (Peek-a-boo system).

Various codes have been devised, each corresponding to one of the three types of representations which have been previously discussed: abstract ornament (cf. 3.1, two-dimensional design), artifacts (cf. 3.2, three-dimensional constructions), paintings, engravings, etc. (cf. 3.3, Iconographical compositions). A brief account will now be given of their general structure, so as to provide a concrete illustration of our method.

4.1. ABSTRACT ORNAMENT

Individual motives, on the one hand, are defined by the combination of a *root,* indicating an elementary sign (d for dash, s for an S, b for the loop, etc.) with one or several *affixes,* indicating various geometrical arrangements (r for radial, li for linear, m for symmetrical, etc.) (see Fig. 1). The distinction between both kinds of morphemes is purely operational; it is based on considerations of economy and substantiated by statistical observations.

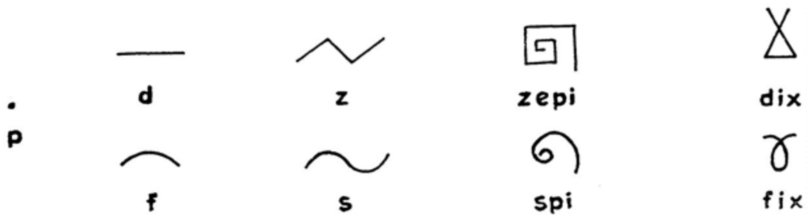


FIGURE 1. *(a)* Elementary signs, examples.

Each elementary sign could actually be regarded as a specific arrangement of one single fundamental sign, the dot. The linguistic system built on that convention would, however, be very unwieldy for the purposes of documentation.

With only twenty elementary signs (roots), actually distributed in two parallel series of ten signs each, rectilinear and curvilinear, and, on the other hand, ten sorts of geometrical arrangements (affixes), subdivided into about thirty variants indicated by vowel gradations for each affix, one determines a first series *of primary ornaments* (about six hundred).

After undergoing a second treatment of the same kind, each of these generates a *secondary ornament* (i.e., a root+two affixes, about 18,000 in number), which in turn can develop into a *ternary sign* (i.e., a root+three affixes; 500,000 in number), etc. At the fourth degree, the number of motives reaches fifteen millions, each of them being named in terms of the very condensed language mentioned above, with less than thirty morphemes (see Fig. 1).

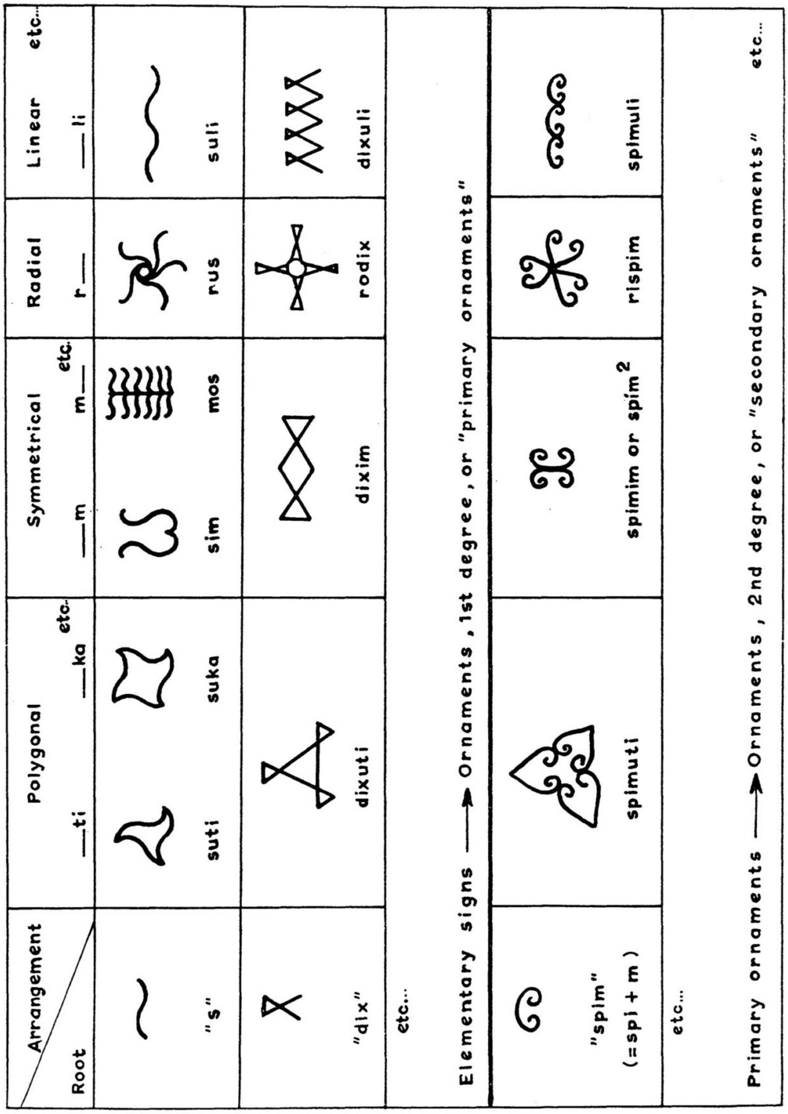


FIGURE 1. (b) Combinations of elementary signs (roots) in various arrangements (affixes).

Through various conventions, one can analyze according to the same notions the most complex ornamental compositions, where several motives of the 1st, 2nd, 3rd, and 4th degrees come into play. The relative position of the various motives is expressed by a third set of terms, most of which are the same as roots and affixes, but are written in block letters to indicate their syntactical function (Fig. 2).

FIGURE 2. (a) "spi F," several spirals (spi), arranged on a curve (F). (b) "suti KA," several "suti" (three S forming a triangle) arranged in square (KA).

In a longer formula, the position which any symbol holds is naturally as important as its absolute meaning when considered separately. Provision has therefore to be made to indicate, through different devices (index numbers, for instance) the interrelations of the various terms in the formula.

4.2. ARTIFACTS

4.21. Five different parts of *pottery shapes* are considered: base, main body, neck, handle, and spout. A series of conventions fixes the dividing lines between base and body, neck and body, etc., so as to channel individual appreciations into a unique pattern, close to common usage. Each part is then de

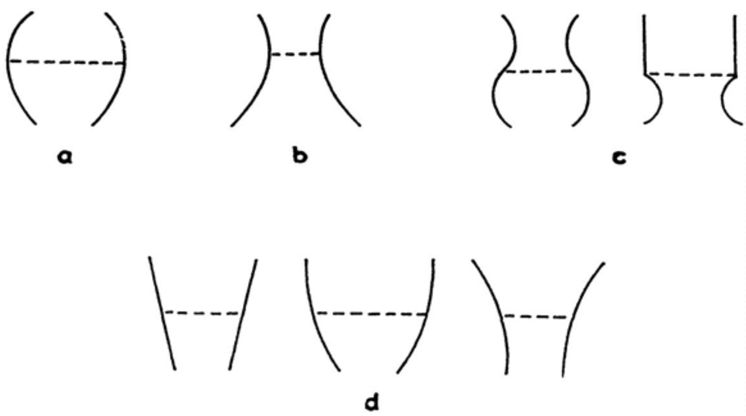


FIGURE 3. Pottery, main body: conventional division between upper part and lower part.

scribed through a combinatorial analysis, which always proceeds in the same way:

(a) Conventional division into various subparts: The main body, for instance, is examined successively under two different cross-sections, horizontal and vertical. In turn, the vertical section is divided into two conventional parts, which it has proved convenient *always* to distinguish on each side of a horizontal plane imagined at one of the following levels (Fig. 3): maximum (a) or minimum (b) width, change of curve (c), or, in all other cases, medium height (d).

(b) Morphological description of each subpart, in terms of a mnemotechnical code such as the following:

d, straight

v, concave

x, convex

i, divergent

u, parallel

o, convergent

Figure 4 shows how the combinations of the six terms, under two different

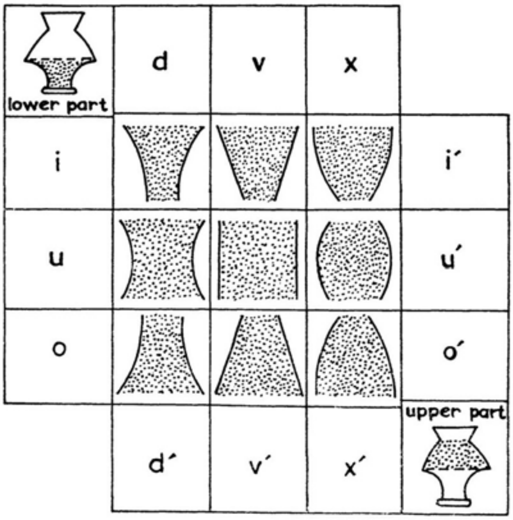


FIGURE 4. Pottery, main body: 9 elementary profiles.

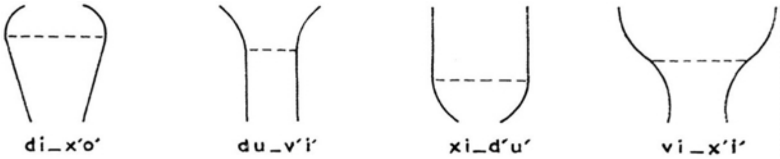


FIGURE 5. Pottery, main body: examples of code words for different shapes.

headings, shape (d,v,x) and relative inclination of the sides (i,u,o), provide nine elementary profiles for each subpart, i.e., 81 shapes for the whole body. (See examples, Fig. 5.)

(c) Description of the liaison between the various subparts: Figure 6 illustrates, for instance, three additional terms (c, curve; 1, angle; q, moulding), which, combined with the 81 previous formulas, determine 243 types of profiles for the main body.

At this stage the purely morphological description of any of those 243 shapes is effected with only five terms, chosen among a total vocabulary of nine.

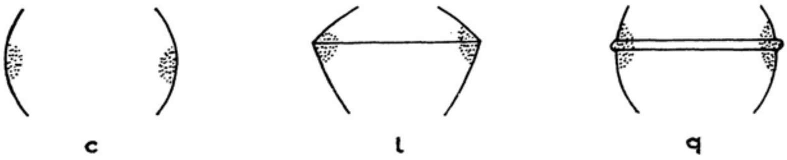


FIGURE 6. Pottery, main body: liaison between upper and lower part.

(d) Relative dimensions: In addition to morphology, the analysis also has to consider dimensions, both absolute (a simple matter of determining successive *bands,* from x to y, for the most significant measures, height, width, etc.) and relative. Figure 7 shows the three ratios which have been taken into consideration in the case of the main body: height of the upper part to height of the lower part, total height to maximum or minimum width, width of the opening (either on the neck or on the base side, according to the morphological type at hand), to maximum or minimum width. Simple fractions or integers are used to determine a suitable scale of variations—five steps for each of the first two ratios, and two for the third one. The delineation of the main body is thus determined to a sufficient degree of approximation by

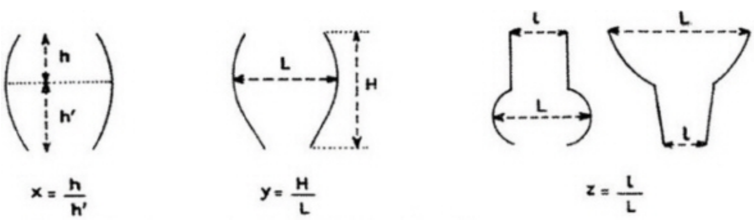


FIGURE 7. Pottery, main body: relative dimensions:

1 x<1/2	6 y<1/2	z<1/2
2 1/2<x<1	7 1/2<y<1	z>1/2
3 x#1	8 y#1	
4 1<x<2	9 1<y<2	
5 x>2	0 y>2	

adding three dimensional terms to the five morphological terms previously mentioned. The number of shapes which can be differentiated in this way amounts to 12,150, with the use of eight terms in each case, chosen among a total of only 27. Actually, among those 27 terms, eleven recur twice (d,v,x; i,u,o; and each of the five arithmetical expressions), so that the total number of descriptive features amounts in fact to only sixteen distributed under six headings.

The procedure is the same for analyzing other parts or subparts of a pottery; the foregoing features of description recur several times in different contexts. For the whole analysis the proposed code does not include more than about thirty terms, each of which corresponds to a simple notion which can be expressed without ambiguity in any language.

4.22. *Tools and weapons.* A combinatorial method has been developed in the same fashion to describe metal implements. In fact, only the preliminary *carving up* (i.e., the establishing of conventional divisions) differs; for each part and subpart, the successive steps of analysis and the terms which are used in the process remain unchanged, whether the artifact be a container, a sickle, or a spear.

4.3. ICONOGRAPHICAL MONUMENTS

Two codes have been constituted, one for Greek coins, where iconographical themes are relatively simple, and the other for oriental seals, where, on the contrary, the number of elements brought into play is very high. Others are being developed for other fields, such as Egyptian sculpture and Indian reliefs; however, they all exhibit the same general structure, which can be summarized in the following way:

4.31. First comes a *vocabulary,* in which the various nominal features of representation are distributed under ten main headings:

- Animate beings, with their individual attitudes (posture, gestures) and angle of representation
    - Humans (i.e., human-shaped)
    - Animals
    - Hybrids
- Inanimates
    - Clothing
    - Building and furniture
    - Containers
    - Instruments (tools, weapons, etc.)
    - Emblems
    - Nature (sky, earth, plants)
    - Ornament

A first kind of coding takes place within each of the ten sections. Instead of developing a formal analysis which would prove unwieldy, it is more convenient to treat the various elements in a given class (houses, axes, stars, etc.) as entities, each of which is to be defined by a *drawing,* and named by a conventional symbol (a figure or a letter) placed after the general designation (Ex.: house, type b; axe, type M7).

Ambiguities or long periphrases are thereby avoided. Moreover, whenever two elements are found to be mutually exclusive in a given iconographical field, their respective variations can be specified with the same set of symbols, as illustrated in Fig. 8. For example, in Greek numismatics, *temples* and *houses* are never found together on the same coin; similarly, *chariots* and *boats* are mutually exclusive, etc. That procedure has two merits: it reduces the number of terms in the code, and it provides a level of generalization (temples, houses, etc.), where phenomenological subtleties can be provisionally ignored, if the scope of a given investigation makes it desirable.

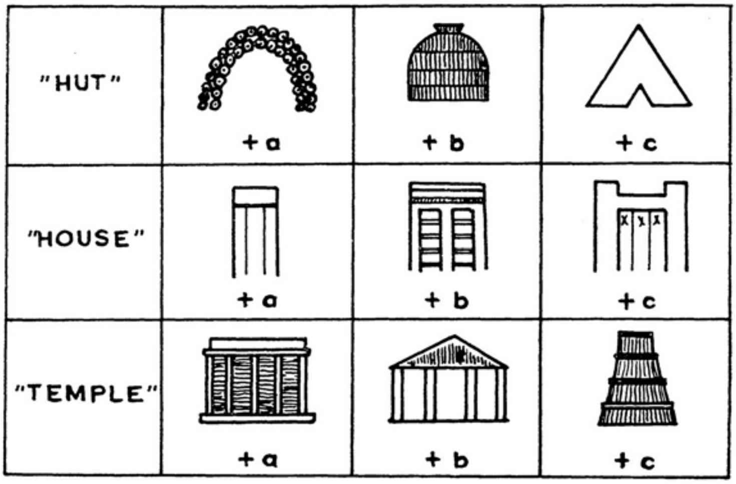


FIGURE 8. Variants of three different motives, designated with the same set of symbols a, b, c, etc.

4.32. Another section in the code deals with the *logical relations* which are explicit in the picture.

Actions. The manifold variety of actions represented on the iconographical documents which we have considered has been analyzed in terms of only two notions, negative and positive, specific meanings being provided by the context.

The *negative* refers to actions which are detrimental to the object, whereas the *positive* embraces all other actions, whether they be beneficient from the point of view of the object, or simply *neutral* That simple dichotomy is generally all that is needed to differentiate iconographical themes such as the following:

man, ewe, +positive action=to tend, to shear, to feed, to milk, etc.

_____ +negative action=to slaughter, to immolate, to sacrifice, etc.

man, building, +positive action=to build, to repair, etc.

_____ +negative action=to demolish, to plunder, etc.

There remains, admittedly, some ambiguity; but the meaning is specified by other terms in the formula:

man, ewe, positive action+container=(probably) to milk

______ "+instrument=______ to shear

_____ "+plants=______ to fodder

Where this is not so, the ambiguity generally comes from the representation itself, not from the way in which it has been analyzed: the killing of a lamb may be related to magic (divination), religion (sacrifice), alimentation (slaughter), etc., without any change in the purely graphical expression of the event, from one case to the other. The lack of precision is here an accepted consequence of the objectivity which must be achieved in the coding process.

Cases. There remains however another kind of ambiguity, due to the different roles which a given element may play in the picture (see *logical relations* in Section 3.32). That problem is solved by affecting to each nominal term a *grammatical case,* chosen among the following:

subject	Ex. A coin adorned with a *tree*.
object	Ex. A man cutting a *tree*.
qualificative	Ex. A man seated, holding a *branch*.
instrumental	Ex. A hero killing a lion with a *branch*.
locative	Ex. A woman seated in a *tree*.

Various logical devices, supported by statistical observations, make it possible to abstract those grammatical features themselves, for certain classes of iconographical elements (humans, animals, plants, etc.), so as not to raise unduly the amount of terms in the code. In this way, the ratio of the number of inflected terms to the number of corresponding radicals can be reduced to a low value, usually smaller than two.

4.33. Iconographical codes are thus constituted with a relatively small number of combinable elements, chosen at different levels of analysis: nominal components, with their morphological variants, grammatical cases, and verbal relations. None of those codes exceeds for the present a thousand terms; yet, the number of representations which can be defined with such terms is very large, from simple motives to the most complex iconographical themes, with a high degree of concision and objectivity, and at any level of generalization which may be required in the course of research.

5. GENERAL REMARKS

It is likely that the method which has been outlined is adaptable to the analysis of other kinds of graphical representations. One might, for instance, consider its application to the tabulation of crystallographic and organic bodies, for documentation purposes in mineralogy and chemistry. The indexing of trademarks, of standardized but variegated pieces of equipment in specific technical fields could also be opened to similar procedures. Photographic or schematic recordings of data yet “unexplained” in terms of a scientific theory could perhaps also be indexed in the same way.

The sole condition is that such representations should exhibit a certain degree of homogeneity, or orderliness. A code can then be proposed, which is but one of several interpretations of the observed order.

Formulations of that kind, undertaken for *practical* purposes (the storage and retrieval of information) are not guided by the care for logical elegance only. Like all linguistic systems, they tend to establish a compromise between the conflicting needs of economy and rapidity of communication. They are not directly concerned either with the meaning of the various notions to which they lead, from a scientific point of view; their main purpose is to substitute combinations of well-defined terms for synthetical descriptions which usually convey information either too vague or insufficient.

Subject-Word Letter Frequencies with Applications to Superimposed Coding

HERBERT OHLMAN

ABSTRACT. The frequencies of occurrence of English letters in the first five positions of subject words and proper names are determined. With these frequencies a superimposed code is designed. No code book is required. Coding space is utilized almost as economically as with a random code. An empirical check is made. A quantitive measure of word popularity is proposed using letter-frequency data.

Coding, or the transforming of information from one guise to another, is one of man's commonest activities. Every picture may be said to be a coding of some real scene and every written word a coding of some utterance—the brain itself is said to work with coded impulses.

Since the beginning of mass communications, starting with the invention of printing, and increasing with the widespread use of electronics, efficient use of existing space and time has become more and more important. Today, information theory provides a sound basis for determining the limits of transmission speed and accuracy. However, Shannon's theory (1) does not tell us how to make a particular code more efficient. The design of codes is still an art; this paper deals with the improving of one particular type, superimposed coding. In information searching, mechanical aids are being used wherever possible. For a machine to process information, the information must be coded, usually into some variant of that most basic code of all, the binary. However, the most efficient code for pure selection appears to be a superimposed random code. Each coding position is used in a random manner, and a group of coding positions contain superimposed entries.

Calvin Mooers (2, 4) calls such coding "Zatocoding" and has applied it in his patented marginal-punched card system called Zator. However, Zatocoding requires an intermediate step in both coding and searching—a code

HERBERT OHLMAN System Development Corp., Santa Monica, Calif.

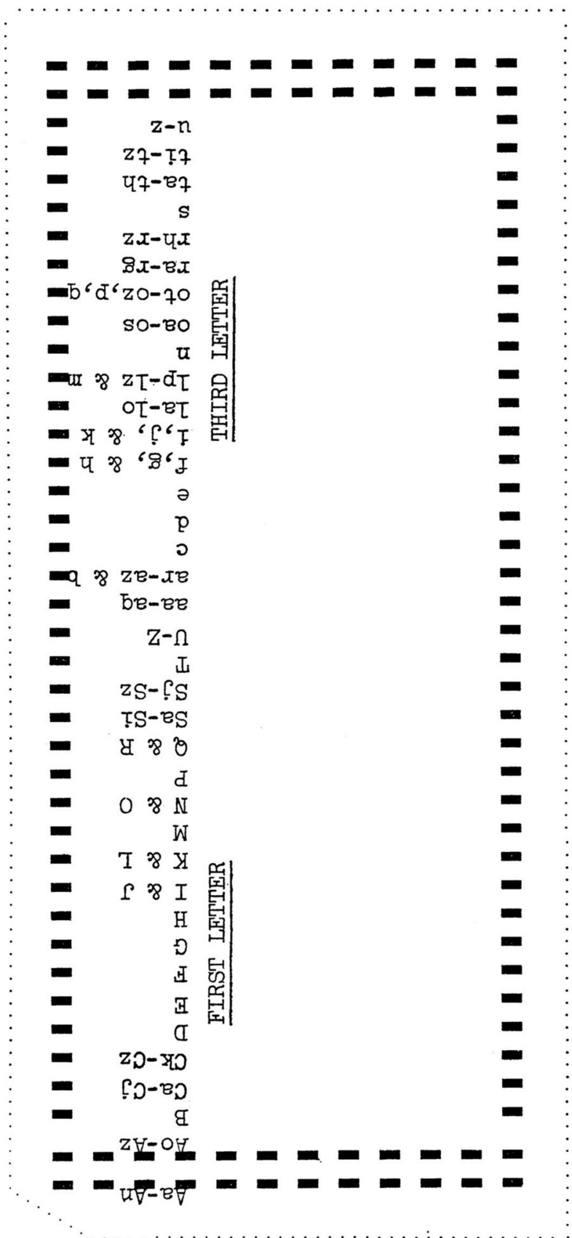

FIRST LETTER
Aa-An
Ao-Az
B
Ca-Cj
Ck-Cz
D
E
F
G
H
I & J
K & L
M
N & O
P
Q & R
Sa-Si
Sj-Sz
T
U-Z
THIRD LETTER
aa-aq
ar-az & b
c
d
e
f,g, & h
i,j, & k
la-lo
lp-lz & m
n
oa-os
ot-oz,p,q
ra-rg
rh-rz
s
ta-th
ti-tz
u-z

FIGURE 1.

book containing a number of indexing terms with random-number equivalents.

Carl Wise (3,4) has produced a nonrandom superimposed code which he calls "word coding" for use with marginal-punched cards of the Keysort variety. This type of code does not require an intermediate code book.

The author has attempted to combine the best features of both systems in coding English words by essentially pre-randomizing the alphabet. This is possible because there is a certain invariance of the letter frequencies within each letter position of a word.

As this system was developed in response to a specific need, it may be well to talk about it in concrete terms, and later apply its principles to other information systems. A marginal-punched card produced on IBM equipment (5) was used as the unit record. The thirty-eight positions along the top edge could code 38 words or phrases by using a direct code, but by superimposition every position could be made to do multiple duty. However, neither of the two systems previously described seemed to meet the requirement of a directly interpretable, yet efficient code.

Subject-word and proper-name lists were studied to find what letter frequencies occurred in the first five letter positions. Some work along these lines had been done, notably by Geisler for the ASM-SLA (6) with proper names, and by Krieger (7) with subject words (however, Krieger only considered initial letters in designing his code).

Striking similarities for initial-letter frequencies among various subject-word lists were found, as shown in Table 1. The average of five such lists show that 40% of the words begin with C, S, P, or A (in that order). Furthermore, 85% begin either with these four or B, M, T, R, E, F, D, G, H, or I—or only 54% of the alphabet.

Even greater consistency was found with proper-name lists, as shown in Table 2, but with a different ranking of the letters. The average of three such lists gave S, B, M, H, and C for the beginning letters of 40% of the names, and these five and D, G, K, L, R, P, W, A, and F (again 54% of the alphabet) accounted for 83%.

The Library of Congress list was chosen as typical of the subject-word lists, and the 1955 Syracuse Telephone Directory as typical of names. A systematic sample was obtained from each list by recording the top-left, middle, and top-right terms from every two-page spread.[1] The frequencies of letters in each of the first five positions were then obtained for each list, as shown in Tables 3 and 4.

[1] The probabilities in this case are not independent, but every term is equidistant in the alphabetical sequence from the next term chosen, which is a sufficient approximation to true randomness for the purposes of this study.

TABLE 1 *Subject-word initial letter frequencies*[a]

Letter	Chamber's Technical Dict'y, 1942, 912 pp.		Merriam-Webster Unabridged Dict'y, 2987 pp.		Industrial Arts Index (Vol. 41, No. 5), April, 1953, 787 pp.		Chem. Abstracts Decennial Subject Index 1907–16 & 1927–36 (after Krieger, ref. 7)		Lib. of Congress T.I.D. List of Subject Headings, June 1952, 327 pp.		Average Frequency, %	Rank
	Freq., %	Deviation	Freq., %	Deviation	Freq., %	Deviation	Freq., %	Deviation	Freq., %	Deviation		
A	7.3	−1.0	*6.6*	−1.7	9.3	+1.0	8.6	+0.3	*9.7*	+1.4	8.3	4
B	6.2	+0.4	5.7	−0.1	*4.5*	−1.3	*7.4*	+1.6	4.9	−0.9	5.8	5
C	10.8	+0.4	9.8	−.06	10.0	−0.4	*12.6*	+2.2	*8.8*	−1.6	10.4	1
D	*5.7*	+1.5	4.9	+0.7	3.3	−0.9	3.5	−0.7	3.6	−0.6	4.2	11
E	4.7	+0.3	*3.3*	−1.1	*6.1*	+1.7	3.9	−0.5	4.2	−0.2	4.4	9
F	4.6	+0.3	3.9	−0.4	3.7	−0.6	4.9	+0.6	4.5	+0.2	4.3	10
G	3.7	−0.2	3.2	−0.7	4.6	+0.7	3.7	−0.2	4.5	+0.6	3.9	12
H	4.4	+0.6	3.7	−0.1	3.3	−0.5	4.3	+0.5	3.6	−0.2	3.8	13.5
I	3.1	−0.7	3.0	−0.8	*6.0*	+2.2	3.7	−0.1	3.2	−0.6	3.8	13.5
J	0.6	0.0	0.9	+0.3	0.2	−0.4	~0.4	−0.2	1.7	+0.1	0.6	21.5
K	1.0	+0.4	0.9	+0.3	0.4	−0.2	~0.4	−0.2	0.3	−0.3	0.6	21.5
L	3.8	+0.7	3.2	+0.2	2.5	−0.6	3.3	+0.2	2.6	−0.5	3.1	15
M	5.6	−0.1	5.0	−0.7	6.6	+0.9	5.3	−0.4	6.2	+0.5	5.7	6
N	2.0	−0.5	1.8	−0.7	2.4	−0.1	3.1	+0.6	3.2	+0.7	2.5	16
O	2.2	0.0	2.6	+0.4	1.7	−0.5	2.7	+0.5	1.9	−0.3	2.2	18
P	9.1	−0.4	9.3	−0.2	10.3	+0.8	*10.8*	+1.3	*8.1*	−1.4	9.5	3
Q	0.6	+0.1	0.6	−0.1	0.2	−0.3	~0.7	+0.2	0.3	−0.2	0.5	23
R	4.4	−0.4	4.8	0.0	4.7	−0.1	*3.5*	−1.3	*6.8*	+2.0	4.8	8
S	10.4	+0.3	*12.4*	+2.3	9.9	−0.2	*8.4*	−1.7	10.4	+0.3	10.1	2
T	4.8	−0.5	6.3	+1.0	5.1	−0.2	4.3	−1.0	6.2	+0.9	5.3	7
U	0.8	−0.4	1.9	+0.7	0.9	−0.3	~0.7	−0.5	1.9	+0.7	1.2	20
V	1.6	+0.2	1.7	+0.3	1.1	−0.3	1.4	0.0	1.3	−0.1	1.4	19
W	1.8	−0.5	3.3	+1.0	2.4	+0.1	1.9	−0.4	1.9	−0.4	2.3	17
X	0.2	0.0	0.1	−0.1	0.2	0.0	~0.03	−0.2	0.3	+0.1	0.2	26
Y	0.2	−0.1	0.4	+0.1	0.1	−0.2	~0.03	−0.3	0.3	0.0	0.3	24.5
Z	0.4	+0.1	0.3	0.0	0.2	−0.1	~0.03	−0.3	0.3	0.0	0.3	24.5
Check sum	100.0	+0.5	99.6	0.0	99.7	+0.2	99.6	0.0	99.7	+0.2	99.5	

[a] Frequencies which deviate more than 1% from the average are shown in italics.

TABLE 2 *Proper-name initial letter frequencies* [a]

Letter	*Chemical Abstracts Fourth Decennial Author Index, 3531 pp.*		*ASM-SLA Metal. Literature Study, 4870 pp.*		*Syracuse, N. Y., Telephone Directory, 1955, 307 pp.*		*Av. Freq., %*	*Rank*
	Freq., %	*Deviation*	*Freq., %*	*Deviation*	*Freq., %*	*Deviation*		
A	3.8	0.0	*2.45*	−1.35	*5.2*	+1.4	3.8	13
B	9.2	−0.2	10.2	+0.8	8.8	−0.6	9.4	2
C	5.7	−0.9	6.2	−0.4	*7.8*	+1.2	6.6	5
D	5.0	−0.3	5.3	0.0	5.5	+0.2	5.3	6.5
E	2.3	+0.1	2.25	+0.05	2.0	−0.2	2.2	16
F	3.6	0.0	3.4	−0.2	3.9	+0.3	3.6	14
G	5.3	0.0	5.6	+0.3	4.9	−0.4	5.3	6.5
H	6.7	−0.1	7.35	+0.55	6.2	−0.6	6.8	4
I	1.6	+0.7	0.75	−0.15	0.3	−0.6	0.9	21.5
J	1.8	+0.1	1.75	+0.05	1.6	−0.1	1.7	19
K	*6.0*	+1.2	4.9	−0.3	4.6	−0.6	5.2	8
L	4.6	−0.4	5.65	+0.65	4.6	−0.4	5.0	9
M	7.7	−0.6	8.25	−0.05	8.8	+0.5	8.3	3
N	2.3	+0.3	1.8	−0.2	2.0	0.0	2.0	17
O	1.4	−0.2	1.4	−0.2	2.0	+0.4	1.6	20
P	4.6	−0.1	4.5	−0.2	4.9	+0.2	4.7	11.5
Q	0.1	−0.1	0.1	−0.1	0.3	+0.1	0.2	25
R	5.0	+0.1	4.65	−0.25	4.9	0.0	4.9	10
S	11.3	+0.1	11.0	−0.2	11.4	+0.2	11.2	1
T	3.4	+0.2	3.65	+0.45	2.6	−0.6	3.2	15
U	0.7	+0.2	0.45	−0.05	0.3	−0.2	0.5	23
V	1.8	−0.1	2.15	+0.25	1.6	−0.3	1.9	18
W	4.6	−0.1	4.65	−0.15	4.9	+0.2	4.7	11.5
X	0.0	0.0	0.0	0.0	0.0	0.0	0.0	26
Y	0.5	+0.1	0.5	+0.1	0.3	−0.1	0.4	24
Z	1.0	+0.1	1.1	+0.2	0.7	−0.2	0.9	21.5
Check sum	100.0	+0.1	99.55	−0.4	100.1	−0.2	100.3	

[a] Frequencies which deviate more than 1% from the average are shown in italics.

TABLE 3 *Subject-word letter frequencies (332 words)*[a]

	First letter		Second letter		Third letter		Fourth letter		Fifth letter	
Letter	Freq., %	Rank	Freq., %	Rank	Freq., %	Rank	Freq., %	Rank	Freq., %	Rank
A	9.3	2	17.8	1	8.4	2	7.7	3	5.3	9.5
B	4.8	8	0.6	17	2.7	15	2.2	17	0.0	24.5
C	8.1	3	1.8	12	5.4	9	6.2	5	2.3	11.5
D	3.6	12.5	0.3	21.5	6.3	6.5	5.0	9	2.0	13
E	4.2	11	12.3	2	6.3	6.5	11.8	1	13.3	1
F	4.5	9.5	0.3	21.5	2.4	16	1.2	20	0.7	19
G	4.5	9.5	0.3	21.5	1.8	17	0.9	22	0.7	19
H	3.6	12.5	3.9	9.5	0.9	19	3.7	13.5	1.3	15
I	3.3	14.5	11.1	4	5.2	10	10.8	2	9.3	4
J	0.9	21	0.0	25.5	0.0	25.5	1.2	20	0.0	24.5
K	0.6	23.5	0.3	21.5	0.3	22.5	2.8	15.5	0.3	21.5
L	2.7	16	6.9	7	7.8	5	3.9	11.5	6.7	7
M	6.3	6	0.9	14.5	3.9	12.5	5.3	8	2.3	11.5
N	3.3	14.5	3.9	9.5	5.7	8	5.9	7	7.3	6
O	2.1	17.5	11.4	3	8.1	3.5	6.2	5	11.7	3
P	7.8	4	1.5	13	3.9	12.5	3.7	13.5	1.0	16.5
Q	0.6	23.5	0.3	21.5	0.3	22.5	0.3	24	0.0	24.5
R	6.6	5	7.5	6	12.0	1	3.9	11.5	12.0	2
S	9.9	1	0.6	17	4.5	11	4.3	10	5.7	8
T	6.0	7	2.4	11	8.1	3.5	6.2	5	8.7	5
U	1.8	19	7.8	5	3.3	14	2.8	15.5	5.3	9.5
V	1.5	20	0.3	21.5	0.6	20	1.2	20	0.3	21.5
W	2.1	17.5	0.0	25.5	0.3	22.5	0.3	24	0.7	19
X	0.6	23.5	0.9	14.5	0.3	22.5	0.0	26	1.0	16.5
Y	0.3	26	6.0	8	1.2	18	1.9	18	1.7	14
Z	0.6	23.5	0.6	17	0.0	25.5	0.3	24	0.0	24.5
No. of blanks	0		0		0		9		33	
Check sum	99.6		99.7		99.7		99.5		99.6	

[a] Blanks are not counted in computing percentages.

TABLE 4 *Proper name letter frequencies (309 names)*[a]

Letter	First letter Freq., %	First letter Rank	Second letter Freq., %	Second letter Rank	Third letter Freq., %	Third letter Rank	Fourth letter Freq., %	Fourth letter Rank	Fifth letter Freq., %	Fifth letter Rank
A	3.2	14	21.7	1	6.8	5	7.0	5.5	6.5	8.5
B	8.8	3	0.3	19	2.3	14	4.3	10	1.5	16
C	7.8	4	3.9	7	2.9	12	3.6	12.5	2.9	11.5
D	5.5	6	0.6	15.5	2.0	16.5	6.0	8	1.5	16
E	2.0	17	14.3	2	7.1	4	11.6	1	19.2	1
F	4.2	13	0.0	23.5	1.3	20	0.7	23	0.7	20
G	4.9	9	0.6	15.5	3.9	10	2.6	14.5	2.9	11.5
H	6.2	5	3.6	8	2.3	14	2.6	14.5	6.5	8.5
I	0.6	22	10.0	4	4.9	8	8.0	3	6.9	6
J	1.6	19	0.0	23.5	0.3	25	0.0	25.5	0.4	23
K	4.5	11.5	0.0	23.5	1.3	20	3.6	12.5	3.3	10
L	4.5	11.5	3.2	9.5	11.0	3	9.3	2	7.6	3
M	9.4	2	1.6	13	0.6	23.5	1.3	19.5	2.2	14
N	2.0	17	2.3	12	12.0	2	7.6	4	6.9	6
O	2.0	17	16.8	3	6.5	6	4.3	10	10.1	2
P	4.9	9	0.6	15.5	1.6	18	1.7	18	0.7	20
Q	0.3	24.5	0.0	23.5	0.0	26	0.0	25.5	0.0	25
R	4.9	9	6.5	6	12.6	1	6.3	7	7.2	4
S	11.7	1	0.3	19	4.9	8	4.3	10	6.9	6
T	2.9	15	3.2	9.5	4.9	8	7.0	5.5	2.5	13
U	0.6	22	6.8	5	3.6	11	2.3	16.5	0.7	20
V	1.3	20	0.3	19	1.0	22	2.3	16.5	0.0	25
W	5.2	7	0.6	15.5	2.3	14	1.0	21.5	1.5	16
X	0.0	26	0.0	23.5	0.6	23.5	0.3	24	0.0	25
Y	0.3	24.5	2.6	11	2.0	16.5	1.3	19.5	0.7	20
Z	0.6	22	0.0	23.5	1.3	20	1.0	21.5	0.7	20
No. of blanks	0		0		0		6		32	
Check sum	99.9		99.8		100.0		100.0		100.0	

[a] Blanks are not counted in computing percentages.

TABLE 5 *Amount of information (H) in subject-word letters*[a]

Rank, n	First letter	p_n	$-p_n\log_2 p_n$	Second letter	p_n	$-p_n\log_2 p_n$	Third letter	p_n	$-p_n\log_2 p_n$	Fourth letter	p_n	$-p_n\log_2 p_n$	Fifth letter	p_n	$-p_n\log_2 p_n$	Av. English text (after Pratt (9))	p_n	$-p_n\log_2 p_n$
1	S	.099	.3303	A	.178	.4432	R	.120	.3671	E	.118	.3638	E	.133	.3871	E	.131	.3841
2	A	.093	.3187	E	.123	.3719	A	.084	.3002	I	.108	.3468	R	.120	.3671	T	.105	.3414
3	C	.081	.2937	O	.114	.3571	O,T	.081	.2937	A	.077	.2848	O	.117	.3622	A	.082	.2959
4	P	.078	.2871	I	.111	.3520			.2937			.2487	I	.093	.3187	O	.080	.2915
5	R	.066	.2588	U	.078	.2871	L	.078	.2871	C,O,T	.062	.2487	T	.087	.3065	N	.071	.2709
6	M	.063	.2513	R	.075	.2803	D,E	.063	.2513			.2487	N	.073	.2756	R	.068	.2637
7	T	.060	.2435	L	.069	.2661			.2513	N	.059	.2409	L	.067	.2613	I	.063	.2513
8	B	.048	.2103	Y	.060	.2435	N	.057	.2356	M	.053	.2246	S	.057	.2356	S	.061	.2461
9	F,G	.045	.2013	H,N	.039	.1825	C	.054	.2274	D	.050	.2161	A,U	.053	.2246	H	.053	.2246
10			.2013			.1825	I	.052	.2218	S	.043	.1952			.2246	D	.038	.1793
11	E	.042	.1921	T	.024	.1291	S	.045	.2013	L,R	.039	.1825	C,M	.023	.1252	L	.034	.1659
12	D,H	.036	.1727	C	.018	.1043	M,P	.039	.1825			.1825			.1252	F	.029	.1481
13			.1727	P	.015	.0909			.1825			.1760	D	.020	.1129	C	.028	.1444
14	I,N	.033	.1624	M,X	.009	.0612	U	.033	.1624	H,P	.037	.1760	Y	.017	.0999	M,U	.025	.1330
15			.1624			.0612	B	.027	.1407	K,U	.028	.1444	H	.013	.0815			.1330
16	L	.027	.1407			.0443	F	.024	.1291			.1444	P,X	.010	.0664			.1129
17	O,W	.021	.1170	B,S,Z	.006	.0443	G	.018	.1043	B	.022	.1211			.0664	G,Y,P	.020	.1129
18			.1170			.0443	Y	.012	.0766	Y	.019	.1086			.0501			.1129
19	U	.018	.1043			.0251	H	.009	.0612			.0766	F,G,W	.007	.0501	W	.015	.0909
20	V	.015	.0909			.0251	V	.006	.0443	F,J,V	.012	.0766			.0501	B	.014	.0862
21	J	.009	.0612	D,F,G, K,Q,V	.003	.0251			.0251			.0766	K,V	.003	.0251	V	.009	.0612
22			.0443			.0251	K,Q,W, X	.003	.0251	G	.009	.0612			.0251	K	.004	.0319
23			.0443			.0251			.0251			.0251				X	.002	.0179
24	K,Q,X,Z	.006	.0443			.0251			.0251	Q,W,Z	.003	.0251	B,J,Q,Z	0				.0100
25			.0443	J,W	0		J,Z	0				.0251				J,Q,Z	.001	.0100
26	Y	.003	.0251							X	0							.0100
$H=\sum_{n=1}^{26}-p_n\log_2 p_n$			4.2920			3.7964			4.1255			4.5201			3.8413			4.1300
$\log_2 26 = 4.7$																		
$R = 1-H/(\log_2 26)$			9%			20%			12%			4%			18%			12%

[a] Average of five letters, 20.5753/5 = 4.1151.

TABLE 6 *Subject-word cumulative letter frequencies (in rank order)*[a]

Rank	First letter: Letter	First letter: Σ %	Second letter: Letter	Second letter: Σ %	Third letter: Letter	Third letter: Σ %	Fourth letter: Letter	Fourth letter: Σ %	Fifth letter: Letter	Fifth letter: Σ %	Av. English text (after Pratt (9)): Letter	Av. English text (after Pratt (9)): Σ %
1	S	9.9	A	17.8	R	12.0	E	11.8	E	13.3	E	13.1
2	A	19.2	E	30.1	A	20.4	I	22.6	R	25.3	T	23.6
3	C	27.3	O	41.5	O, T	28.5	A	30.3	O	37.0	A	31.8
4	P	35.1	I	52.6		36.6	C, O, T	36.5	I	46.3	O	39.8
5	R	41.7	U	60.4	L	44.4		42.7	T	55.0	N	46.9
6	M	48.0	R	67.9	D, E	50.7		48.9	N	62.3	R	53.7
7	T	54.0	L	74.8		57.0	N	54.8	L	69.0	I	60.1
8	B	58.8	Y	80.8	N	62.7	M	60.1	S	74.7	S	66.2
9	F, G	63.3	H, N	84.7	C	68.1	D	65.1	A, U	80.0	H	71.5
10		67.8		88.6	I	73.3	S	69.4		85.3	D	75.3
11	E	72.0	T	91.0	S	77.8	L,R	73.3	C, M	87.6	L	78.7
12	D, H	75.6	C	92.8	M, P	81.7		77.2		89.9	F	81.6
13		79.2	P	94.3		85.6	H, P	80.9	D	91.9	C	84.4
14	I, N	82.5	M, X	95.2	U	88.9		84.6	Y	93.6	M, U	86.9
15		85.8		96.1	B	91.6	K, U	87.4	H	94.9		89.4
16	L	88.5	B, S, Z	96.7	F	94.0		90.2	P, X	95.9	G, Y, P	91.4
17	O, W	90.6		97.3	G	95.8	B	92.4		96.9		93.4
18		92.7		97.9	Y	97.0	Y	94.3	F, G, W	97.6		95.4
19	U	94.5	D, F, G, K, Q, V	98.2	H	97.9	F, J, V	95.5		98.3	W	96.9
20	V	96.0		98.5	V	98.5		96.7		99.0	B	98.3
21	J	96.9		98.8	K, Q, W,X	98.8		97.9	K, V	99.3	V	99.2
22	K,Q,X,Z	97.5		99.1		99.1	G	98.8		99.6	K	99.6
23		98.1		99.4		99.4	Q, W,Z	99.1	B,J,Q,Z	—	X	99.8
24		98.7		99.7		99.7		99.4		—	J	99.9
25		99.3	J, W	—	J, Z	—		99.7		—	Q	100.0
26	Y	99.6		—		—	X	—		—	Z	100.1

[a] On an equiprobable basis, each letter would occur 3.846% of the time.

TABLE 7 *Weighted letter frequencies, %*[a]

Letter	First letter	Third letter	Fourth letter	Letter	First letter	Third letter	Fourth letter
A	8.5	8.2	7.6	O	2.2	7.9	6.0
B	5.3	2.7	2.5	P	7.4	3.6	3.5
C	8.0	5.1	5.9	Q	0.6	0.3	0.3
D	3.8	5.8	5.1	R	6.4	12.1	4.2
E	3.9	6.4	11.8	S	10.0	5.0	4.3
F	4.4	2.2	1.1	T	5.6	7.7	6.3
G	4.6	2.1	1.1	U	1.7	3.3	2.7
H	3.9	1.1	3.6	V	1.5	0.7	1.3
I	3.0	5.2	10.5	W	2.5	0.5	0.4
J	1.0	0.0	1.1	X	0.5	0.3	0.0
K	1.1	0.4	2.9	Y	0.3	1.3	1.8
L	2.9	8.2	4.6	Z	0.6	0.2	0.4
M	6.7	3.5	4.8	Check sum	99.5%	100.3%	99.9%
N	3.1	6.5	6.1				

[a] All seven parts subject plus one part name.

For the initial letters of subject terms, the rank order was S, A, C, P, R, M, T, · · ·; for second letters, A, E, O, I, U, R, L · · ·; for third, R, A, O or T, L, D or E, · · ·; for fourth, E, I, A, T or O or C, N, · · · ; and for fifth, E, R, O, I, T, N, L · · ·, as shown in Table 5. Cumulated frequencies are given in Table 6.

Table 5 also gives the information measure $-p_n \log_2 p_n$ for each letter in each position (8). For this purpose, percentage frequencies were assumed to represent actual probabilities, p_n. The sum for each letter position,

$$\sum_{n-1}^{26} = p_n \log_2 p_n$$

represents *H,* the average uncertainty per letter-position or, as it is sometimes called, the average information represented by the letter position, in bits. The redundancy *R* is also shown on the bottom for each letter position.

These calculations show that the least redundant (or the most informative) letter position is the fourth, next to that the first, and then the third. Similar results can be shown for proper names.

For the marginal-punched card application, first and third letter positions were selected for coding. Subject-word frequencies were weighted with proper names in a 7-to-1 proportion,[2] as shown in Table 7. The 52 letters of

$$G = H - H\frac{(H-1)^X}{H} = 6.8$$

A maximum of 8 coding words was chosen, based on these calculations.

[2] According to Wise (3), the ratio X/H, or that of the number of positions to be punched to the number of positions available for punching, should be about 0.46. Taking H to be 19, X=8.75. The dropping fraction $f_d=(G/H)^Y$, or the ratio of the number of positions actually punched to the number available for punching, raised to a power, *Y*, representing the number of sorting elements used, works out to be $(7/19)^2$ =13.7%, if about 9 codes are actually superimposed. Note that

first and third positions were then assigned to the 38 available positions as equally as possible, but under the restriction that alphabetical order along the side of the card be preserved. The result is shown in Fig. 1, and Table 8 shows the predicted frequency distribution for this code. Note that it was necessary sometimes to combine several letters in one position, and sometimes to split

TABLE 8 *Comparison of actual and predicted letter frequencies*

First letter[a]

Letter	*Actual no. of cards dropped*	*Actual %*	*Predicted %*
Aa–An (median between anx and any)	29	2.6	4.25
Ao–Az	19	1.7	4.25
B	54	4.9	5.3
Ca–Ci (median between cka and cke)	65	5.9	4.0
Ck–Cz	103	9.35	4.0
D	46	4.2	3.8
E	37	3.35	3.9
F	36[c]	3.3	4.4
G			4.6
H			3.9
I & J			4.0
K & L			4.0
M			6.7
N & O			5.3
P			7.4
Q & R			7.0
Sa–Si (median between siv and six)			5.0
Sj–Sz			5.0
T			5.6
U–Z			7.1

Third letter[b]

Letter	*Actual no. of cards dropped*	*Actual %*	*Predicted %*
aa–aq (median between ard and are)	40	3.8	5.45
ar–az & b	80	7.6	5.45
c	60	5.65	5.1
d	55	5.2	5.8
e	100	9.45	6.4
f, g & h	80	7.6	5.4
i, j & k	45	4.25	5.6
la–lo (median between lov and low)	45	4.25	5.85
lp–lz & m	80	7.6	5.85
n	80	7.6	6.5
oa–os (median between otf and oth)	45	4.25	5.9
ot–oz, p, q	35	3.3	5.9
ra–rg (median between rge and rgo)	40	3.8	6.05
rh–rz	65	6.1	6.05
s	65	6.1	5.0
ta–th (median between tid and tie)	40	3.8	3.85
ti–tz	35	3.3	3.85
u–z	70	6.6	6.3
Total	1060		
Avg.	59		

[a] Ideally each first letter position would comprise 5%.
[b] Ideally each third letter position would comprise 5.5%.
[c] Not carried to completion.

Note: About 400 cards were used in study. Actual number was estimated by measuring cards dropped at 150 cards/inch. Predicted percentage based on Table 7.

Dropping fraction, $F_d = (G/H)^Y = (2.7/18)^1 = 15\%$. For 400 cards, $F_d = 60$. H is the number of coding positions; G is the number of punches / card = 1100/400; Y is the number of sorting positions = 1. (See Wise (3) for derivation.)

one letter between two positions. These splits were chosen according to the median frequencies of English trigrams (9).

Splitting letters modifies the first-letter position somewhat by the second, and third somewhat by the fourth. Such letter-pair frequencies take account of intersymbol influence, and therefore make possible a better code than single-letter frequencies. H.P.Luhn has designed a superimposed code using randomizing squares (10) which takes advantage of letter-pair frequencies.

An empirical check of the letter code shown in Fig. 1 was made on a 400-card file maintained by the author. The results are shown in Table 8[3]. The average dropping fraction for the third position alone compares well with the dropping fraction as calculated by formula, but the range (from 9 to 25%) is broader than hoped for. However, Table 8 shows that the agreement between actual and predicted frequencies in the third-letter position was very good, considering the alphabetic-order limitation imposed in assigning the positions.

By using data-processing equipment, much more elaborate studies on much larger samples would be possible. The author is working with such equipment and hopes to have some results available in the near future.

Equifrequency-letter codes have many other applications, including the preassignment of space in files and indexes, in cryptography, and in philology. For example, the data in Table 5 can provide a quantitative measure of subject word popularity. Taking a few words from the Library of Congress list of subject headings, we add the percentage frequencies of each letter (up to 5) together and divide by the number of letters. (Multiply each p_n by 100 to get the percentage frequency.)

AIRCRAFT has a value of 9.3+11.1+12.0+6.2+12.0, 10.12

DIVIDER has a value of 3.6+11.1+0.6+10.8+2.0, 5.66

ICHTHYOLOGY has a value of 3.3+1.8+0.9+6.2+1.3, 2.70

These three words give some idea of the range possible in a subject-heading list. In general dictionary words, the highest found was SARI, with a value of 12.63, and the lowest, ONYX, with a value of 1.8. It is interesting to compare these values with the highest possible letter combination (not necessarily an English word), which is SAREE (value 12.96), and the lowest (value 0.06), The highest is very nearly realized in actuality, while the lowest never comes close. Also note that the word SARI is certainly uncommon English; this phenomenon may occur because the intersymbol connections are broken by taking single-letter frequencies.

[3] Since the first-letter positions showed quite wide deviations from the predicted frequencies, their analysis was never completed. It is now thought that third and fourth letter positions would have made a more invariant code, less subject to the fluctuations which occur in any particular file because of the selection of particular terms.

ACKNOWLEDGMENT

The work described in this paper was performed while the author was in the employ of Carrier Corporation, Syracuse, New York.

REFERENCES

1. C.E.SHANNON, A Mathematical Theory of Communication, *Bell System Technical Journal,* July 1948, and following.
2. C.N.MOOERS, Zatocoding and Development in Information Retrieval, *ASLIB Proc.,* February 1956, p. 3 (Many other papers by this author may be obtained from his Zator Co., 79 Milk Street, Boston, Massachusetts.)
3. C.S.WISE, A Punched-Card File Based on Word Coding, pp. 93–114, in Perry and Casey's *Punched Cards,* Reinhold Publishing Corporation, New York, 1951.
4. MOOERS and WISE had discussions in *American Documentation,* April 1950, October 1950, and October 1952.
5. H.OHLMAN, The Low-Cost Production of Marginal-Punched Cards on Accounting Machines, pp. 123–26 *American Documentation,* April 1957.
6. JOINT COMMITTEE OF ASM AND SLA, ASM-SLA Metallurgical Literature Classification, American Society for Metals, 1950. (Figure 5, which was based on an analysis of 4870 names by A.H.Geisler in *ASM Review of Metallurgical Literature.)*
7. K.A.KRIEGER, A Punched-Card System for Chemical Literature, *J. of Chemical Education,* March 1949, p. 163.
8. E.T.KLEMMER, Tables for Computing Informational Measures, p. 75 in Quastler's *Information Theory in Psychology,* Free Press, Glencoe, Ill., 1955.
9. F.PRATT, Secret and Urgent, The Story of Codes and Ciphers, Blue Ribbon Books, Garden City, N.Y., 1942, pp. 264–78.
10. H.P.LUHN, Superimposed Coding With the Aid of Randomizing Squares for Use in Mechanical Information Searching Systems, IBM Product Development Lab., Poughkeepsie, New York, 1956.

The Analogy between Mechanical Translation and Library Retrieval

M.MASTERMAN, R.M.NEEDHAM, and K.SPÄRCK JONES

ABSTRACT. Any analogy made between library retrieval and mechanical translation is usually made by assimilating library retrieval to mechanical translation. We desire to draw the converse analogy; that is, to assimilate mechanical translation to library retrieval. To do this, mechanical translation procedures must be generalised and made interlingual, until they become as general as library retrieval procedures already are. This generalisation can be made if the mechanical translation procedure is based on a thesaurus. The nature of a thesaurus is discussed in Section 3. This type of procedure has already been used for library retrieval, but not for M.T.; the use of a thesaurus for both fields enables a new, very general field to be exactly defined, namely the field of semantic transformation. This field would have application to library retrieval, mechanical translation, and probably also to mechanical abstracting. The purpose of this paper is to develop the application of this generalized procedure to mechanical translation, referring also to its use for library retrieval. For this purpose, an analytic examination of the translation procedure is required, as linguists object to the analogy that we are making by asserting that a library retrieval type of procedure will not translate syntax.

It is asserted that a generalised mechanical translation procedure cannot translate grammar and syntax as these do not correspond between different languages. There is a general answer, and a particular one, to this criticism. The general answer is that present procedures for translating between different pairs of languages generate such complexity that they do not form an adequate basis for future M.T. research. The experimental work done by workers in the U.S.S.R. is examined. The particular answer is that since recent mechanical translation experiments using a thesaurus show, contrary to expectation, that this method can interlingually translate semantic meaning, it seems not impossible that, again contrary to expectation, it can be used to translate syntax.

Such an extension is suggested by the linguist M.A.K.Halliday. He defines the syntactic operators of a source language in terms of a set of interlingual *questions*. This procedure is criticised.

M.MASTERMAN, R.M.NEEDHAM, and K.SPÄRCK JONES Cambridge Language Research Unit, Cambridge, England.

A generalised translation procedure, using a thesaurus, is related to the semantic problems of mechanical translation. A thesaurus is defined. Recent work done by the Cambridge Language Research Unit is described to illustrate this procedure. Experiments, done also in the C.L.R.U., using the same procedure for library retrieval, are described. The result, a conception of a procedure of generalised semantic transformation, is considered.

This semantic transformation procedure is extended to cover syntax. The questions used by Halliday can be turned into thesaurus heads. Some examples of interlingual translation of syntactic form are given. Research on these lines is continuing. If this method of generalised mechanical translation proves feasible, M.T. becomes straightforwardly an extended case of generalised retrieval.

PROPOSAL TO CREATE A SINGLE GENERAL THEORETIC FIELD OF SEMANTIC TRANSFORMATION, WITH APPLICATION TO LIBRARY RETRIEVAL AND TO M.T.

Many documentalists have insisted that there is an analogy between mechanisable procedures for retrieving documents and procedures used in mechanical translation (M.T.). The analogy between the two has usually been drawn, however, by assimilating library retrieval to translation; not the other way round. A coded library classification has been envisaged as an exact and interlingual library language. Any request for information, made in a particular language, must be translated into the interlingua, and also coded, if the retrieval procedure is to be mechanical (1).

We wish to draw the analogy conversely: that is, by assimilating interlingual mechanical translation to retrieval. Now, in the present state of research this analogy can only be drawn at all precisely between one form of library retrieval procedure, and one form of mechanical translation procedure; these two analogous procedures are those, in each field, which make use of a thesaurus. The proposal that an improved type of library retrieval procedure could be devised, using a thesaurus, of the type of Roget's famous *Thesaurus,* instead of a term classification, has already been made by American workers in this field (2,3). The proposal that semantic meaning can be translated using a thesaurus was first made by the Cambridge Language Research Unit (England), at the Second International Conference for Machine Translation (4,5,6). We propose, then, that a conceptually based, thesaurus type of language classification should be used for a completely generalised retrieval procedure, this classification procedure being, by its nature, interlingual. The development of this procedure makes possible the definition of a general theoretic field of semantic transformation. Of this field, a well-defined mathematical model can be made (7).

Surprisingly enough, the proposal that such a general field should be created seems far more revolutionary to mechanical translation specialists than to documentalists specialising in library retrieval. Translation specialists, and, in particular, linguists deny even the possibility of the analogy by maintaining that any classification of language based on a thesaurus can, at best, only hope to translate semantic meaning, whereas language is primarily a system of grammar and syntax; and both of these are notoriously monolingual. It could be said, indeed, that a library classification is like a non-grammatical language and that a thesauric library retrieval procedure could therefore hope to retrieve from it. But it is obvious, so the argument runs, that any mechanical translation procedure, before it starts dealing with subtle questions of semantic ambiguity, must deal with crude questions of how to translate grammatical and syntactic form; and these are both notoriously monolingual. Since, therefore, grammar and syntax cannot be translated by an interlingual thesaurus procedure, the analogy we wish to draw falls to the ground: it has no application to any procedure for mechanical translation.

The object of this paper is to refute this criticism by showing how a type of retrieval procedure, based on a thesaurus already being used for the experimental translation of semantic meaning, might also be extended so as to translate grammar and syntax. It is only by showing the procedure in action that we can hope to make clear what seems to us this most fundamental and important analogy between library retrieval and mechanical translation; we hope to show the nature of the generalised procedure by considering how it can deal with the particular problems of one of the fields in question, namely M.T. And this is all the more necessary in that the field of mechanical translation, unlike that of library retrieval, has not hitherto been approached at all from this point of view.

1. APPLICATION OF THE METHOD TO M.T.

On July 10th, 1957, M.A.K.Halliday read a paper, to the Cambridge Language Research Unit, and later in a developed form, to the International Congress of Linguists, held in August, 1957, in Oslo, in which, speaking as a descriptive linguist, he described a method which might be used to carry out an interlingual analysis of the syntax of a language (8,9). This method was nicknamed the Twenty Questions Method of Analysis.

Before discussing the method, however, we must give a provisional reply to those M.T. workers who deny the existence of an analogy between the mechanical translation and retrieval fields. These may ask, “Why attempt an interlingual translation between languages when we know that the grammar and syntax of different languages do not correspond?" They may also ask, "Since it is mechanical translation of technical material which is urgently required in order to make scientific information more generally available, why not have, as the U.S.S.R. mechanical translation workers have, a set of two-language programmes, to translate from, e.g., Italian into French, or from Chinese into Russian, using for any particular text the appropriate programme?"

The answer to these questions, still keeping for the moment within the M.T. field, is that those who use such an approach, constructing a separate programme, to be stored by the machine, for every pair of languages, fail to consider the complexity which the method itself generates. Only one group of workers has extensively tried this method out: the Mechanical Translation Research Group of the U.S.S.R. Academy of Sciences. The project is described in an informative recent paper by I.K.Belskaya (10). This paper explicitly sets forth the restrictions on translation necessary to limit the complexities generated by the method itself. These are (1) *severe limitation of the input text:* only mathematical texts were used, the translation being from Russian into English; and the U.S.S.R. group only at present envisages mechanical translation of scientific texts; (2) *limitation of vocabulary:* in order to limit the number of multiple meanings required for successful dictionary entries, a separate entry was used for each whole word—the attempt to economise on storage space by dividing words into "chunks," or sub-words (11) was abandoned; (3) *multiplication of dictionaries:* different dictionaries were required for all the different fields, even when translating between the same pair of languages.

These experiments show that a mechanical translation programme constructed on the Russian model does not straightforwardly translate between two languages. What such a translation programme does, when used with, e.g., a technical mathematical dictionary and a general dictionary containing the common words of the language, is successfully to translate English mathematical texts into Russian. This is a tremendous technical achievement. But it is inadequate as a directive for future research. The failures, cited by Belskaya, of attempts by cryptographers and logicians to find a common basis, statistical or mathematical, to language, might indeed cause us to abandon the goal of interlingual translation. But we cannot abandon the attempt to achieve intertextual translation. If we cannot feed into a computer and translate, from a single source language, e.g., a novel, a philosophical treatise, a mathematical system and a botanical paper, without using separate programmes and dictionaries, we are not translating between pairs of languages. We are merely translating between pairs of texts. And mechanical translation on this basis is not a commercial prospect.

If we reconsider the Russian experiments, therefore, with the necessity for intertextual translation in mind, we are tempted to ask, "Can we at once have a more general approach to the problem?" This question seems all the more appropriate when we find that the U.S.S.R. group themselves think that a more general attempt to translate syntax might be successful. Belskaya says:

Special experiments were made in order to find out whether the same grammatical programme can be applied to a text having as little to do with mathematics as, say, an article from *The Times,* or a page from Charles Dickens. These experiments proved the success of our ideas on the possibility of having a universal grammatical programme for the machine translation of any two languages. Our general principles have withstood another test: they were extended to cover machine translation from languages differing from English in structure as much as Japanese, Chinese, and German. These experiments having been successful, the principles (underlying the Russian grammar and syntax programme) may be considered as basic in the solution of machine translation problems.

Thus even the U.S.S.R. group, whose approach is strictly particularised and inductive, admit that there may be general ascertainable principles underlying the mechanical translation of grammar and syntax. The next object, then, of linguists associated with machine translation, ought to be the discovery and development of these principles, rather than further experiments on particular texts. We propose that this research should be pursued by substituting for the particularised methods of linguistic analysis at present in use among workers on M.T. the completely generalised methods at present in use in library retrieval; that these, having been given thesauric linguistic application, should be put on a machine, and the results examined. Such a method, which is essentially algorithmic and deductive, does not, of course, invalidate the step-by-step method of inductive generalisation, at present being used in U.S.S.R. But the light that it throws upon the whole process of semantic transformation, and the simplifications which can be attained by means of it, make it in our view a preferable basis for the next stage of research.

2. A SUGGESTED INTERLINGUAL ANALYSIS OF SYNTAX

That M.T. research could be thus generalised is the opinion already of one linguist, M.A.K.Halliday. We must next, therefore, examine and criticise the method he suggests for the interlingual mechanical translation of grammar and syntax, before further considering the problem of whether fully interlingual and intertextual mechanical translation of scientific texts is possible.

Halliday's method was first to make a strictly monolingual analysis of the input language. He then made a further interlingual analysis of the language. For this interlingual analysis he does not recommend a generalised transfer grammar, of the kind developed by the American descriptive linguists, Z.Harris and N.Chomsky (12,13). He recommends using a more direct analytic method. This owes much to 19th century historical linguists. But Halliday's analysis, unlike theirs, is not evolutionary. First, he makes a rigid distinction between types of chunk, the operators of a language, and the arguments. (Roughly, the functions of operators are dealt with by grammar books; those of arguments, by dictionaries.) The operators are identified by their relation, positive or negative, to a number of categories (provisionally about 60). The arguments are then classified by referring to groupings of these systems (14).

Basically, therefore, Halliday makes first a monolingual grammar, and then an interlingual analysis of each language, the latter being quite distinct from the former. The monolingual grammar resembles those of descriptive linguists, except that it refers only to operators; the arguments are later defined by referring to the operators. The interlingual analysis, the key to the whole method, demands reference to extralinguistic contexts; only after these have been ascertained are the operators related to the arguments. The relation of any operator to the extralinguistic context is determined by asking questions, the answer to which can be “Yes,” “No,” “Both,” “Neither.” This procedure resembles that of the game “Twenty Questions,” from which the method derives its name. The two methods differ, however, in that, for the linguistic analysis, in most cases, the answer to one question does not influence the next.

The interlingual analysis may proceed as follows. Take, for example, the French operator *la*. A normal grammatical description would classify this as either the feminine definite article, or the feminine accusative pronoun. We assume that *la* has already been subjected to a monolingual French analysis giving, e.g., gender. We now carry out the interlingual analysis: we do not ask “Does *la* belong to any gender system?” because it is notorious that the gender systems of different languages do not correspond. Therefore we simply ask: “Can *la* tell us anything about sex?” By this change of question we refer, not to the intralinguistic context (i.e., that of French), but to the far more general extralinguistic context (i.e., that of the human race divided into sexes). English has no genders, French has two, German three, Icelandic six, but English, French, Germans, and Icelanders alike fall into communities of only two sexes. Therefore the answer to our last question is “Yes.” We may then ask: “Does *la* refer to animate or inanimate objects?” The answer is “Both.” To the question “Does it apply to present or non-present time?” the answer is “Neither.” And so on.

Now it is clear that, even from the pure linguist's point of view, Halliday's suggestion is of great research interest, since what he proposes is to use the precise and elegant analytic methods of contemporary linguistics to analyse, both monolingually and interlingually, the context grammar of particular texts. (These analytic methods, as is known, depend on being able to break up the older grammatical units, such as noun, verb and the rest, into weaker but more precisely definable units, special to each language, from which, by referring to the intralinguistic context-grammar of a text, the older type of unit, can, where it is required for that particular language's analysis, be built up.) In order to extend this method to make it apply to an interlingual grammar based on extralinguistic context analysis, it is evident that Halliday must take seriously the analogy, to which older linguists have paid nothing more than lip service, between intralinguistic context and extralinguistic context, and the way that each might be used to build up grammar and syntax. And, from the pure linguistic point of view, this is a very interesting thing to do. But if we consider his interlingual analysis from the point of view of mechanical translation rather than from that of linguistics, it is clear that it has serious defects. These are (1) that the monolingual analysis is too complicated a way of obtaining the list of operators of an input language; a first approximation to these could be obtained with far less trouble by consulting a grammar book, and then, by applying the procedures, to find out where and why the translation had turned out wrong; (2) that though the method *analyses,* it does not *translate*. For mechanical translation purposes it must be turned from a method of analysis into a translating procedure; (3) that the method is essentially not linguistic at all, but logical. Therefore logical sophistication, rather than linguistic scholarship, should be used to make the question system more economical.

3. A PROCEDURE FOR THE TRANSLATION OF SEMANTIC MEANING, USING A THESAURUS

This bringing to bear of logical methods on problems of M.T. is at present being tackled by only one unit. Only the Cambridge Language Research Unit uses logical methods together with linguistics for mechanical translation research. It is no coincidence, therefore, that it is the only unit which is simultaneously investigating procedures for mechanical translation, library retrieval, and mechanical abstracting. Although it is almost universally assumed, by mechanical translation research groups, that it is the linguist, not the mathematician, who provides the computer programmer with the data for the translation programme, we contest this. We consider that the very nature of the problem of interlingual mechanical translation is like that of information retrieval in that it demands a general, that is, a logical approach. Belskaya considers that as no logical system for interlingual translation has yet been devised, none could exist. We hold not only that such a system can exist, but also that it does exist as soon as the output language is analysed, not as a dictionary but as a thesaurus. The interlingual system required for mechanical translation and library retrieval alike is thus not a new interlingual language. It consists of a logical system giving the structural principle on which all languages are based. This principle is that language, seen interlingually, consists of an ordered finite set of clusters of synonyms (the synonyms being, of course, different for each language), which can be represented by a corresponding ordered set of topics, or very general abstract nouns, or heads. These heads are homogeneous, that is, they do not themselves divide up into different parts of speech, since the synonyms of which they represent the common principle of synonymy will be in different parts of speech in different languages. They are vague; their “meanings” cannot be given except by reference to the sets of synonyms in any language which represent them, and these sets of synonyms are not precisely bounded. They are unobservable; some existing word in any given language may be usable, in an extended sense, to represent some idea common to a set of synonyms, or it may not, in which case, either a new word has to be invented, or the set has to be left identifiable only by reference to its position in the total ordered set of heads. Thus, in English, the head-words “greatness,” “smallness,” “region,” “base,” “land,” do exist; the head-word “materiality” does not. In short, these new interlingual units of semantic transformation have a series of theoretic “ineffable qualities” attached to them just as Newton's infinitesimal operators seemed to have to his contemporaries in 17th century philosophy and science. But like Newton's operators, these units also can be used in determinate mathematical procedures; such a procedure is given in detail, for mechanical translation, in Appendix 1, and for library retrieval, in Appendix 2.

Thus the nature of a thesaurus, or general ordering principle for language, may be briefly characterised as follows. A thesaurus, or synonym dictionary, e.g., *Roget's Thesaurus,* unlike an ordinary dictionary, consists of an ordered set of lists of synonyms grouped under a comparatively small number of concepts, or topics, or heads. (We use the word “head” to describe these because it is the word which Roget himself used.) These heads are themselves arranged in a single or multiple hierarchy, usually in decreasing order of generality; thus the chapter of contents of a thesaurus, taken by itself, will exemplify the mathematical system of a tree. The whole thesaurus cannot be taken as a tree, however; because, in it, the words of a language will always occur more than once; a synonym occurring under any given thesaurus head represents only the use of that word in that language in that context. Thus the occurrence of the English word "plant" in *Roget's Thesaurus* under "Agriculture" signifies that "plant," used agriculturally, means "something growing." The occurrence of the same word "plant," under "tool," signifies that "plant" here means, "total engineering apparatus"; and so on. The head-words of the thesaurus do not reoccur; but the synonym words given under them do; and this means that the total thesaurus, consisting of chapter of contents, numbered list of heads, and lists of synonyms, cannot, mathematically speaking, be represented as a tree, but must be represented as a lattice; that is, as a partially ordered set of which any two elements will not only have a point in common above them, higher in the hierarchy; but also a point in common below them, lower in the hierarchy. The advantages for library retrieval of substituting a lattice for a tree are exemplified in Appendix 2; the advantages for mechanical translation are exemplified in Appendix 1 by the fact that the mechanical translation procedure does obtain an output. Thus the theoretic importance of making the new field of semantic transformation work on a lattice, rather than on a tree is that the lattice, unlike the tree, guarantees that translation-points, and retrieval points, in the system do exist; that is, that, by using the system, information can be retrieved, and translations obtained.

In practise, both for library retrieval and for translation, it is so important to be able to locate the end points of the semantic transformation procedure that the thesaurus itself is always used in conjunction with a cross-reference dictionary. In *Roget's Thesaurus* the thesaurus itself occupies only the first half of the book; all the second half is occupied by a cross-reference dictionary in which those words of the English language which occur in the thesaurus are listed alphabetically, each word being followed by a list of the numbers of the thesaurus heads in which the word occurs. Now in *Roget's Thesaurus* itself the cross-reference dictionary is of course unilingual; *Roget's Thesaurus* is normally used only by authors for improving their style, that is, for translating from English into English. When interlingual translation is contemplated, however, the cross-reference dictionary, for any source language, must be bilingual; that is, its use must transform the chunks of any input text into sequences of lists of thesaurus heads (see Appendix 1). Thus, although interlingual translation is contemplated, for each source language there must be a separate cross-reference dictionary. The lattice of thesaurus heads will be interlingual; but the lists of synonyms, idioms, and paraphrases appearing under any given head will be individual to any output language. Thus each output language must have its own lists of synonyms to fit under the heads of the interlingual thesaurus. And the process of transforming the sequence of source-units (terms in the request, in the library case, chunks in the source text, in the translation case) into sequences of sets of thesaurus heads; of operating an algorithm to select from among these heads and/or to substitute for them other heads or sets of heads from the total thesaurus; and of transforming the selected sets of heads into output (documents or sub-documents, in the library case, synonyms in common between the selected heads, in the translation case), this total process constitutes the process of semantic transformation, and the total possible field to which it can apply the proposed new general theoretic field.

It is worth remarking that, when this procedure is used, translation, like retrieval, becomes irreversible and asymmetrical. The words of the source language must be divided into as fine sub-words as possible, so that the dictionary entry of each chunk shall give a whole list of head-numbers; this list, if it is full enough, exactly defines the spread of that chunk's ambiguity, and distinguishes this spread from the ambiguity-spreads of cognate chunks. Analogously, in the library case, the library-user's request, if it is at all complex, must be analysed as finely as possible into terms. On the other hand, the synonym lists, in the output language, must be as long, complete, and vivid as possible, consisting not only of whole words, but also of whole phrases, sometimes even of whole sentences. Analogously, in the library case, the output consists of a series of whole documents; or, at the least, of references to sections or paragraphs within them.

To recapitulate: a thesaurus, as Roget himself saw (15), is not primarily an analytic tool; it gives a procedure for finding analogies; that is, for finding relevant information; that is, for finding translations. It is organised on a hierarchical principle which is like that of a library-retrieval tree classification, except that, instead of forming the mathematical system of which the model is a tree, it forms the mathematical system whose model is a lattice. It is interlingual, in the sense that the heads have synonyms in any language. It gives a general solution for semantic problems; that is, those arising from the unusual use of words or from multiple meaning. It deals therefore, and by a single procedure, with the most difficult problems facing alike mechanical translation research, and research into methods of mechanising information retrieval, and methods of mechanising the process of sub-titling and abstracting. Of course, information retrieval is a perfect field for applying a thesaurus procedure, just because a library system may be regarded as a language without syntax. But we claim to have shown also that the use of a thesaurus has immense possibilities also for mechanical translation itself. In experiments performed at the C.L.R.U., a thesaurus procedure has been used *(a)* to translate a novel use of a word in an Italian scientific paper, *(b)* mechanically to translate a line of Latin poetry, (c) to retrieve documents from a library, and *(d)* mechanically to construct the essentials of discursive paragraphs of text (7). In all of these, the thesaurus has been used only to translate semantic content. So we return once more to the question, "Can it also translate syntax?"

4. THE SAME PROCEDURE, RELATED TO SYNTAX

We already have the method devised by Halliday to analyse syntax. His type of questions, for instance, were used to analyse the Latin sentences, "Magnam multitudinem vidit" and "Ad ludum ambulamus." (The actual questions used were those obtained by M.Masterman and K.Spärck Jones.) We now have to ask, can a thesaurus procedure derived from these be used to translate? For all Halliday's questions can be rephrased as single words; these, in turn, can be replaced by thesaurus heads; and these, by their nature, will yield an English output, when the thesaurus-lists are in English. In principle, therefore, the problem of using Halliday-derived heads for translating instead of analysing can be solved, at a stroke, simply by turning his questions into heads. In practise, however, the general problem of interlingually translating syntax can be resolved into two difficulties: (1) Can the information given by the monolingual analysis of an input text (as done, e.g., by Mukhin (16), Richens (17), and Halliday be picked up by a dictionary leading straight to a thesaurus? Can the entries for such queer chunks as –AT– and –US of the Latin OB-STIN-AT-US lead into combinations of thesaurus heads? The answer is that in some cases they can, though we do not know if this is true of all cases. (2) Can a thesaurus really be used to translate grammar as well as syntax? That is, can grammatical form, for example, the subject-predicate relation, be treated as semantic information leading to a thesaurus? The answer to this again is, in some cases, yes.

Whether such entries can be constructed for all features of monolingual grammar, we do not yet know. What is already evident is that answers to these two questions can only be obtained if the principles of monolingual analysis are reconsidered from the fundamentally different viewpoint of a thesaurus-maker; and this means approaching the problem from the viewpoint, no longer of linguistics, but of retrieval. And this is so because a thesaurus is essentially a logical structure, designed to retrieve relevant information from an antecedently constructed complex, namely a thesaurised language or sub-language.

REFERENCES

1. R.A.FAIRTHORNE, *Patterns of Retrieval.* Amer. Doc. 1956.
2. C.M.MOOERS, Information Retrieval on Structured Content. Paper of Third London Symposium on Information Theory, 1956.
3. H.P.LUHN, A Statistical Approach to Mechanised Literature Searching *I.B.M. Journal of Research,* 1956.
4. M.MASTERMAN, *The Potentialities of a Mechanical Thesaurus,* C.L.R.U.
5. A.F.PARKER-RHODES, *An Algebraic Thesaurus,* C.L.R.U.
6. M.A.K.HALLIDAY, Linguistic Basis of the Thesaurus-Type Mechanical Dictionary and Application to English-Preposition Classification. Papers read at Second International Conference on Mechanical Translation, M.I.T. October, 1956. *Abstracts in Mechanical Translation 3,* [2]. See also, for Halliday's paper, *Mechanical Translation, 3,* [3].
7. A.F.PARKER-RHODES and STEPHEN WHELAN, Appendix to *Information Retrieval and the Thesaurus,* by R.M.Needham and T.Joyce; paper presented to Programme Committee, Area 5, *International Conference on Scientific Information.*
8. M.A.K.HALLIDAY, The Linguistics of Mechanical Translation, *Proceedings of the 8th International Congress of Linguists,* Oslo, August, 1957.
9. M.MASTERMAN, Linguistic Problems of Mechanical Translation, to be published in forthcoming issue of *Mechanical Translation.*
10. I.K.BELSKAYA, Machine Translation of Languages, *Research,* Oct. 1957.
11. R.H.RICHENS and M.A.K.HALLIDAY. Word Decomposition for Mechanical Translation; 8th Meeting of Linguists, Georgetown University 1957.
12. Z.N.HARRIS, Transfer Grammar, *International Journal of American Linguists, 20,* [4], 1954.
13. N.CHOMSKY, *Syntactic Structures Ianna Linguarum,* 4'S Gravenhage, 1957.
14. M.A.K.HALLIDAY'S working charts (unpublished).
15. *Roget's Thesaurus,* author's preface, 1931 edition, Longman's.
16. I.S.MUKHIN, *An Experiment on the Mechanical Translation of Languages carried out on the B.E.S.M.,* published by the U.S.S.R. Academy of Sciences, Moscow, 1956.
17. R.H.RICHENS, General Programme for Mechanical Translation between Any Two Languages via an Algebraic Interlingua Paper read at 2nd International Conference on Mechanical Translation M.I.T. Oct. 1956. (Abstract in *Mechanical Translation, 3,* [2].
18. R.M.NEEDHAM and T.JOYCE, The Thesaurus Approach to Information Retrieval, *Am. Document.,* July 1958.

APPENDIX 1

The contents of this appendix are taken from Margaret Masterman's paper "The Potentialities of a Mechanical Thesaurus."

The parts of the paper referred to are concerned with (1) the translation procedure, and the example, the translation of an Italian paragraph, used to show the procedure; (2) the discussion of particular difficulties which arose due to the unusual use of words.

The translation procedure is concerned, not with *free words,* but with *chunks*. A chunk is defined as "the smallest significant language-unit which can exist in more than one context, and which, for practical purposes, it pays to insert as an entry by itself in an M.T. dictionary" e.g., the Italian free word PIANTATORE is broken up into PIANT—AT—ORE. Each chunk, forming an entry in the M.T. dictionary, can have a number of meanings, or uses.

The range of uses of any chunk in a language, can be so envisaged that it forms a tree, the total dictionary entry of the chunk forming the point of origin of the tree. When any such a tree is connected, for translation purposes, with the corresponding tree in another language, the two trees together form a lattice each point of which looks both ways and is itself a translation point. (For an amplified discussion of trees and lattices vide Margaret Masterman "Fans and Heads," and "Outline of a Theory of Language," Work-papers, C.L.R.U.)

A point on a lattice, or a multilingual dictionary article, is analogous to a topic, or *head* in a single language *thesaurus.* "Discussion of this analogy led to the suggestion that a multilingual M.T. programme might be developed (given an imaginary computer of indefinitely expandable size) in which the multilingual dictionary might be replaced by a target language thesaurus."

A brief account of the programme which was developed and the thesaurus using translation experiment which was carried out on an Italian paragraph, is given below. At this point the procedure was only applied to semantic heads; no attempt was made to analyse syntax by the use of a thesaurus.

THE PROGRAMME

In the programme three operations are carried out on the input text which has been broken up into chunks:

1. The chunks are matched with the chunks of a pidgin dictionary, giving a pidgin English output.
2. The pidgin chunks are matched with the relevant entries in the cross reference dictionary of a thesaurus. (*Roget's Thesaurus,* with additions, was used in this test.) This stage gives an output of thesaurus heads relevant to a greater or lesser degree to the final translation output required.
3. A number of operations, restricted by rules, select from the list of thesaurus heads given under the entries in the cross reference dictionary the one which is appropriate. (This is not the final selection, as under each thesaurus head there is a list of synonyms; this problem was not, however, dealt with in this paper.)

A. *Chunking of the Italian passage*

Each chunk was written on a card which was used for the matching process:

LA PRODUZ-ION-E DI VARIET-A DI PIANT-E PRIV-E DI GEMM-EASCELL-ARI O PER-LE-MENO CON GERMOGL-IA SVILUPP-O RIDOTT-O, INTERESS-A DA TEMPO GENET-IST-I ED AGRONOM-I, TAL-E PROBLEM-A SI PRESENT-A PARTICOLAR-MENT-E INTERESS-ANT-E PER ALCUN-E ESSENZ-E FOREST-AL-I E FRUTT-IFER-I, PER LE PIANT-E DI FIBR-A, MA SOPRATUTTO PER IL TABACC-O, IN QUEST-A COLTUR-A E INFATTI IMPOSS-IB-IL-E MECANIZZ-ARE L'-ASPORT-AZION-E DEI GERMOGL-I, ASCELL-ARI, NECESS-ARI-O D'-ALTRA-PARTE PER OTTEN-ERE FOGLI-E DI MIGLIO-E QUALIT-A.

B. Matching of these chunks with English pidgin chunks using the Italian-English pidgin dictionary

The type of entry in the dictionary is as follows:

AL-	… -Y
FIBR-	… FIBRE
I	… THOSE-WHICH-ARE
GENET-	… GENETIC

This matching gave a very pidgin translation. This was improved by using a syntax lattice procedure which gave synthesis routines, and the following pidgin translation was obtained:

THE PRODUCE-MENT OF VARIETY-S OF PLANT-S WITHOUT AXIL-ARY BUD-S, OR AT LEAST WITH SPROUT-S AT REDUCED DEVELOPMENT-S, INTEREST FOR SOME TIME PAST GENETIC-IST-S AND AGRICULTURE-IST-S. SUCH PROBLEM-S SELF-PRESENT PARTICULAR-LY INTEREST-ING FOR SOME FOREST-Y AND FRUIT BEARING ESSENCE-S, FOR THE PLANT-S OF FIBRE-S, BUT ABOVE ALL FOR TOBACCO. IN THIS CULTIVATE-URE IT BE IN FACT IMPOSSIBLE TO MECHANISE REMOVE-MENT OF ALL THE AXIL-ARY SPROUT-S, ON THE OTHER HAND NECESSARY FOR TO OBTAIN LEAF-S OF BETTER QUALITY-S.

This translation obviously fails at some particular points:

ESSENCE-S for ESSENZ-E,

SPROUT-S for GERMOGL-I,

SELF-PRESENT for SI PRESENTA,

strictly also ASCELL- should have been translated by the vernacular ARMPIT-.

These cases were examined in detail, by using the next stages of the procedure: the thesaurus cross-reference dictionary and the thesaurus. The three cases examined are important in that they represent the unusual uses of words:

ESSENZ-E is being used in a new way.

GERMOGL-I is being used technically.

SI PRESENTA is being used idiomatically.

Therefore the pidgin output must be retranslated.

Roget's Thesaurus was used in the normal way, i.e., the chunk of word in question is looked up in the cross-reference dictionary, e.g., *bud* gives:

(a) bud, head no. 367
(b) beginning 66,* 129*
(c) germ 153
(d) ornament 847*
(e) expand 194
(f) graft
(g) -from 154
(h) -dy 711 890

Cross-references which are asterisked are additions made to Roget so that it is made multilingual, i.e., to ensure that there is a reference corresponding to each chunk of the input text when turned into pidgin. These additions were legitimately made by comparing the synonyms given under related heads to discover those in common. It was necessary as in some cases the list of cross-references given in the dictionary was inadequate, and the thesaurus can only be properly used if a chain of meanings can be followed through the thesaurus. (It is sometimes possible to obtain the required meaning by following up the synonyms given under the heads without making any addition to the list of cross-references.)

C. *Particular cases examined using the thesaurus technique which represents the next stage of the procedure.*

(i) ESSENZ-E … ESSENCE-S

If the chunks FOREST AND FRUIT-BEARING ESSENCE-S are matched with the chunks in the cross-reference dictionary of the thesaurus, the following output is obtained:

forest Head no.	*57, 367, 890
and Head no.	37, 38
fruit Head no.	result 164
	produce 161
	food 298
	profit 775
	forbidden- 615
	reap the -s 973
	-tree 367
	fruitful 168
	fruition 101
	fruitless 169, 645, 732
bearing Head no.	relation 9
	support 215
	direction 278
	meaning 516
	demeanour 692
	-rein 752
	*fruit- 168, 637, 367
	*child- 161
essence Head no. essential	5, 398
	intrinsic 5
	*meaning 516
	great 31
	required 630
	important 642
	essentially 3, 5
	essential stuff 5

Operation 1

Pick out all the numbers which occur more than once in the above output; these are called ring numbers. The following is obtained:

Ring number	*Thesaurus head*	*Source of ring number*
367	Vegetable	Forest, fruit
161	Production	Fruit, bearing
168	Productiveness	Fruit, bearing
516	Meaning	Bearing, essence
5	Intrinsicality	Essence

Operation 2

Pick out all the ring numbers and order in a scale of descending frequency of occurrence. This gives:

5, 367, 161, 168, 516.

Operation 3

Compare, in twos, for common elements, the ring number thesaurus heads which represent the meanings of the pidgin chunks. A comparison, or intersection operation on two thesaurus heads is permitted only if there is some relation between the chunks from which the heads are derived. In this test the relationship is determined by the syntax lattice: an intersection is permitted only if the points on the syntax lattice corresponding to the chunks have an inclusion relation between them. Thus it is permitted to intersect two heads from the same chunk as this is a trivial inclusion relation. We also intersect the heads from different chunks provided there is an inclusion relation between the points representing the chunks on the syntax lattice. The direction of the inclusion relation determines for which of the chunks the output of the head intersection is to be taken as a new translation. The rule is that the output retranslates the lower of the two chunks, i.e., the one included.

In the case of any two chunks, *A* and *B,* the operation of comparison is called $A \cap B$, and we have an ordering as follows:

$$A \cap A = A \equiv A \geqq A$$

A covers *B*

$$A \cap B = B \equiv A \geqq B$$

Carrying out this operation we get an output of which selected examples are given below:

$A \cap A = A \equiv A > A$	*Chunk comparison*	*Ring number comparison (in all possible pairs)*
	Fruit ∩ fruit	367 ∩ 161
	Fruit ∩ fruit	161 ∩ 168
	Fruit ∩ fruit	367 ∩ 168
A covers *B*	Fruit ∩ bearing	161 ∩ 168
	Fruit ∩ bearing	161 ∩ 367
	Fruit ∩ bearing	161 ∩ 516
	Fruit ∩ bearing	168 ∩ 367
	Fruit ∩ bearing	168 ∩ 516
	Fruit ∩ bearing	367 ∩ 516
$A \cap B = B \equiv \geqq B$	Forest ∩ essence	367 ∩ 5
	Forest ∩ essence	367 ∩ 516

Further operations can be carried out by combining the chunks. The comparison Forest ∩ fruit cannot be carried out as their lattice positions are not inclusive; in this case, by chance, no intersections are prohibited, as all possible combinations of the numbers are made by other chunks.

Operation 4

List the common elements (or words) given by the intersection of the thesaurus heads; i.e., those words common to the lists of synonyms given for the separate heads, e.g.,

Ring numbers	*Thesaurus heads*	*Output: Words common to the lists of synonyms for the heads*
5 ∩ 161	Intrinsicality ∩ production	Flower, etc., Head no. 22 (prototype)

This gives a new thesaurus head; it is necessary to carry out the intersection procedure, as under Operation 3, using this new head, with results as follows, e.g.,

5 ∩ 22	Intrinsicality ∩ prototype	example, specimen
516 ∩ 22	Prototype ∩ meaning	prototype, example

Operation 5

We wish to obtain alternative translation for a particular chunk, i.e., to select from a list of synonyms for the chunk a more appropriate word. The synonyms are obtained by applying the output words given by Operation 4 to the intersections of Operation 3. So that we get, e.g.,

Fruit ∩ fruit	367 ∩ 161 (Production ∩ vegetable)	flower

Therefore the final output, flower, is a retranslation of fruit. Similarly for ESSENCE-, the example word selected, we get:

Essence ∩ essence	5 ∩ 516 (Intrinsicality ∩ meaning) and then operations using the new head, 22 (prototype)	example

Example is therefore a retranslation of essence.

In the case of operations on two different chunks the synonym refers to the chunk which comes lower in the lattice.

A number of restrictive rules are required to regulate the final output.

(i) referring to Operation 3, $A \cap A{=}A \equiv A \geqq A$. When $A \geqq A$ is taken, and yields $A \cap A$, if A is not a chunk but a ring number, take the output which is identical with the original chunk.

(ii) If the above rule operates reject all other output.

(iii) When selecting the final output, take the longest output first, i.e., if there is a synonym output for Fruit-bearing essences, prefer it to a synonym for Fruit-bearing.

Using these rules we get the final synonyms, as follows:

for FOREST ESSENCE we get forest flower

for FRUIT-BEARING ESSENCE we get fruit-bearing example,

(ii) GERMOGL-I… SPROUT-S

(iii) SI PRESENTA … SELF PRESENTS

Retranslations for these output words were carried through in exactly the same way as for the first example. The third test failed, however, as to retranslate it adequately the syntax of the whole sentence would have to be taken into account, and the syntax lattice was not developed to this point. Nevertheless, the truncated procedure yielded the interesting translation, “strike one as.”

CONCLUSION

A number of conclusions were drawn from this test which indicated that further work on thesauri for translation purposes might be fruitful.

Claims are made for the thesaurus procedure as following:

(a) It is a procedure for giving an idiomatic translation.
(b) It is possible to see where it goes wrong.
(c) The test gives useful information on the construction of a thesaurus; this would assist the making of a thesaurus directed to M.T.
(d) The only dictionaries used are the bilingual pidgin dictionaries. The major lexical emphasis is on the target language thesaurus; and this one thesaurus serves for translation from *all* languages into the target language.
(e) The procedure uses previous M.T. results, which show the efficiency of mechanical pidgin; at the same time further analysis of the input language is possible.

Difficulties of a computer holding a thesaurus might be solved by encoding the thesaurus in the form of lattices, the points of which represent chunks.

A number of modifications have since been made in the procedure developed in this test, although the idea of using a thesaurus has been continued.

1. The preliminary translation into pidgin was abandoned.
2. The syntax lattice, in the form used in the test, was also abandoned.
3. The matching process was revised.

The successive intersection process has been seen to be uneconomical. The point has been discussed in the Unit's papers on Information Retrieval, and in detail in "A Note on a Property of Finite Lattices" by R.M.Needham.

When the original paper was written the problems of analysing syntax were considerable and were thought unconnected with the thesaurus procedure for semantics.

APPENDIX 2

The brief account given below is taken from those parts of "The Thesaurus Approach to Information Retrieval," by R.M.Needham and T.Joyce, which refer particularly to the use of a thesaurus.

The problem of library retrieval is to describe documents so that, for any request in ordinary language, all relevant documents can be retrieved by a simple operation, without losses or irrelevancies.

If a large number of terms are used to describe a document, the existence of synonyms is likely; in a system such as Uniterm no attempt is made to bracket the synonyms, which means that a request will produce only the document described in identical terms and not in synonymous ones. If the existence of synonyms is avoided, by using a small number of exclusive descriptors, the description of a document in terms useful for retrieval is more difficult, also it is equally difficult to relate a request to the descriptions of documents. A further difficulty is that descriptions only list the main terms, and take no account of their relations to one another. The C.L.R.U. experiments being carried out make use of a thesaurus, a procedure through which it is hoped that these difficulties will be avoided, and that a request for a document although not using the same terms as those in the document, will produce that document and others dealing with the same problem, but described in different, though synonymous, terms.

PROCEDURE

1. Term abstracts are made of the documents; the descriptors of a particular document are thus terms taken from it. It is at this point that the problem of synonyms must be solved without making the description procedure too rigid. The solution is arrived at in the next stage of the procedure.
2. The terms are then arranged so that near-synonyms are accommodated. This can be done by introducing a partial ordering relation, in which more specific terms are included in more general ones dealing with the same topic. So that in making a request for *A,* we are given *B* if the relation $A \geqq B$ holds. This makes allowance for loose description and also for structure.

This ordering is in effect making a thesaurus; each term can from this point of view be described as a head, and the inclusion relations of the terms correspond to the general-specific relations of a group of thesaurus heads, if we discuss the thesaurus in the terms of the ideal rather than the actual. Similarly, the parallel does not quite hold if we consider the list of synonyms given under each head in a thesaurus such as Roget, although it could be made to do so. The synonyms given under each Roget head are not ordered in any kind of relationship except that of being synonyms, or rather, near synonyms, for the head. If we consider the retrieval terms, however, this situation does not exist: the ordering relations are much more fully worked out, so that what would be synonyms in the thesaurus appear as terms either including, or more usually, included in, the original term; only synonyms in the strict sense are equal. Usually the near-synonyms appear as subordinate to, or more specific terms than, the main term. In this way the existence of near-synonyms is allowed, so that there is no loss of information; at the same time loosely expressed requests can be made precise, it being possible to treat each term if necessary as independent. The hierarchy constructed also made it possible, in dealing with requests in the retrieval procedure, to obtain a scale of relevance in order to secure the correct output.

The partially ordered set is converted, by including latent elements, into a lattice, making the ordering of the terms more systematic, and also making the actual retrieval procedure carried out on punched cards (for details see the paper) easier.

Linguistic Transformations for Information Retrieval

Z.S.HARRIS

ABSTRACT. This paper discusses the application to information retrieval of a particular relation in linguistic structure, called transformations.[1] The method makes possible the reduction of a text, in particular scientific texts, to a sequence of kernel sentences, which is roughly equivalent in information to the original text. It seems possible to determine the division into kernels in such a way that each kernel will carry about as much information as is likely to be called for independently of the neighboring information in the article. A text may therefore be stored in this form (perhaps omitting, by means of formal criteria, any sections which are unnecessary for retrieval), and its individual kernels may be retrieved separately. Since the carrying out of transformations depends only on the positions of words in a sentence, and not on knowledge of meanings, it seems possible that at least part of this operation can be performed by machine; the more so since the method does not require any judgment about the subject matter, or any coding of the concepts of a particular science.

1. LANGUAGE STRUCTURE

Methods such as are presented here are possible because language can be objectively studied as a part of nature, and it is then found to have an explicitly stateable formal structure. It can be objectively studied if one considers speech and writing not as an expression of the speaker which has particular, introspectively recognized, meanings to the hearer; but rather as a set of events— sound waves or ink marks—which may be said, if we wish, to serve as tool or

Z.S.HARRIS Department of Linguistics, University of Pennsylvania, Philadelphia, Pennsylvania.

[1] Z.S.Harris, Co-occurrence and transformations in linguistic structure, *Language, 33* (1957), 283–340. A discussion of the place of transformations in linguistic theory is included in Noam Chomsky, *Syntactic Structures* (1957). The further method of discourse analysis mentioned below is preliminarily presented in Z.S.Harris, Discourse analysis, *Language, 28* (1952), 1–30. The application to information retrieval has been investigated with the support of the National Science Foundation.

vehicle for communication and expression. This set of events, the many occurrences of speaking and of writing, can be described in terms of a structural model. We can say that each such occurrence is composed of parts (mostly successive in time), and that these parts (words, or word parts, or sounds) are collected into various classes, e.g., the class of words *N* (noun). These classes are defined not by meaning but by the position of the words relative to each other. Any property of language which is stated only in terms of the relative occurrence of physically describable parts is called "formal."

Each language is found to have a formal structure, and the structures of various languages are similar to each other in various respects. In each language we find that we can set up and classify elements called "morphemes" (word parts, such as prefixes, or indecomposable words) in such a way that the various classes of these morphemes occur (in speech and writing) in a regular way relative to each other. The domain of this regularity is called a sentence. We can identify a few sequences of morpheme classes (and sequences of sequences), and show that every sentence is one or another of these sequences, but this only with the aid of recursive formulations.

In this way, the structure of language can be shown to have algebraic properties. This may seem strange to those who expect language to be irregular or full of exceptions. It is true that there are a great many special details if we try to approach anything like a complete description of a language. And it is true that each word has its own particular range of other words with which it occurs in a sentence; for example, a given noun occurs before certain verbs and not others. Nevertheless, the sequence of word classes that constitutes a sentence is definite for a given language, and even the fact that every word has a unique range of other words with which it occurs turns out to provide a basis for regularities in the language structure, as will be seen in 1.1.[2]

1.1. TRANSFORMATIONS

For our present purposes, the only structural feature we have to consider is transformations, and as much of language structure as is required in order to operate with these. We begin by defining a construction as a sequence of n word or morpheme classes in terms of which a sentence or another construction is structurally described. For example, *AN* (as in *charged atom*) is a two-class sequence, and *TN Ving N* (as in *the atoms emitting electrons*) is a three-class

[2] Aside from this there are homonymities in language, i.e., structurally different words or sequences that have the same sounds and may therefore be physically indistinguishable; these are degeneracies of the language structure. And there are ambiguities, i.e., cases where the hearer cannot tell which of several meanings of a word was intended; structural analysis cannot help in this case, except perhaps by such methods as discourse analysis.

sequence (neglecting the T), both of which constitute a construction "noun phrase" which we may write *N*ʹ.[3] *N is A* (as in *Power is evil.*) and *NV* (as in *justice sleeps.*) are two-class sequences, and *NVN* (as in *Denaturation affects rotations.*) or *N is V en by N* (as in *Results were recorded by observers.*) are three-class sequences, all of which constitute sentences.[4]

A construction is satisfied for certain particular members of its n classes if the sequence of those n words (those members of its n classes) actually occur as a case of that construction. For example, *AN* is satisfied for the pairs *(charged, atom), (clean, room), (large, room)* and a great many others. *T N Ving N* is satisfied for the triples *(atom, emit, electron), (atom, contain, electron), (molecule, emit, electron), (worker, work, overtime)* and a great many others. It might seem that not much can be done with this: the satisfaction list for each construction is very large; and it is impossible to say that one or another n-tuple does not satisfy the construction. Thus we cannot say that the noun phrases *charged room* or *atoms evicting overtime,* would not occur somewhere. We can only say that in some sample of the language the pair *(charged, room)* did not occur in the construction *AN,* or that even in a very large sample the triple *(atom, evict, overtime)* did not occur as a sentence construction. Nevertheless, we can obtain a useful result by considering (for a given sample of the language) not the absolute lists of n-tuples that satisfy a given construction, but the similarities among the lists that satisfy different constructions.

If two constructions *R, S* which contain the same n word classes (with possibly additional individual morphemes, e.g., *AN* and *N is A*) are each satisfied for the same list of n-tuples of members of these classes, we call the two constructions transforms of each other:[5] *R*↔*S.* If the satisfaction list for *R* is a proper subset of the satisfaction list for *S,* we say that *R* transforms one-directionally into *S: R*→*S*. For example *AN*→*N is A:* every pair satisfying *AN* can also be obtained in *N is A* (*The atom is charged,* etc.), but there are certain subtypes of *A* which occur in the latter but not the former. Similarly *T* N_1 *Ving*

[3] Marks for positionally defined word classes in English: *N* noun (*bridge, idea,* etc.), *V* verb *(see, erupt,* etc.), *A* adjective (*new, fortuitous,* etc.), *T* article or quantifier (*the, some,* etc.), *D* adverb (*very, A+ly,* etc.), *P* preposition (*of, without,* etc.), *C* conjunction (*and, but,* etc.). *S* indicates a sequence having sentence structure. For *x, y* ranging over these values, *y* is defined by *Xy=Y,* i.e., *y* is something added to a member of class *X* such that the sequence *Xy* is a member of class *Y* (occurs in the positions of *Y*). E.g., *An=N,* and *-ness, -ism* are members of this *n,* as in *largeness; Vn=N,* and *-ment, -ion* are members of this *n,* as in *treatment, construction*.

[4] Sentences may also be described in terms of noun-phrases and comparable constructions, with *N*ʹ occupying the position of *N,* as in *N*ʹ *is A* (e.g., *The atoms emitting electrons are charged.*).

[5] Since the satisfaction lists for each construction are not closed, and vary somewhat with the sample, "same" here can only mean "approximately same." However, it is possible to establish, by means of linguistic eliciting methods or otherwise, whether this sameness holds between two constructions or does not.

$N_2 \leftrightarrow N_1 V N_2$ (the subscripts distinguish the different occurrences or members of classes): every triple found in one construction can also be found in the other (compare the examples above with *The atoms emit electrons; Denaturation affecting rotations*…). And $N_1 V N_2 \leftrightarrow N_2$ *is V en by* N_1 as in *Electrons are emitted by the atoms; Rotations are affected by denaturation*. But $N_1 V N_2 \not\leftrightarrow N_2 V N_1$: both may be satisfied for the ordered triple *(nature, imitate, art),* if we have *Nature imitates art* and *Art imitates nature;* but we are likely to find the ordered triple *(atom, emit, electron)* only in the first (not finding the sentence *Electrons emit atoms*). Another example: The sentence construction N_1, N_2, V (e.g., *The tapestry, a masterpiece, faded*.) is not a transform of $N_1 V N_2$ (we don't find *The tapestry faded a masterpiece*), but it is a transform of the sentence sequence $N_1 V$. N_1 *is* N_2 (as in *The tapestry faded. The tapestry is a masterpiece.*).

1.2. KERNELS

We thus see that many sentence constructions are transforms of other sentence constructions, or of sequences of these, or include subconstructions (such as noun phrases) which are transforms of some sentence construction. Transformation is an equivalence relation among constructions (in particular, sentences) in respect to their satisfaction lists.[6] The transformations thus provide a partition of the set of sentences. Furthermore, it is possible to find a simple algebra of transformations by showing that some transformations are products of others (base transformations): for example, TN_2 *being Ven by* N_1 and $N_1 V N_2$ meet the conditions for being transforms of each other (*The electrons being omitted by atoms* and *Atoms emit electrons*.); but we can say that this is the product of two transformations which have been mentioned above; $N_1 V N_2 \leftrightarrow N_2$ *is Ven by* N_1 (yielding *Electrons are emitted by atoms*.) and $N_1 V\, N_2 \leftrightarrow T\, N_1$ *Ving* N_2 (where the new V is *is,* and the original *Ven by* is carried after *is* as a constant of the transformation; i.e., *is emitted by*↔*being emitted by*).

Every sentence structure is the product of one or more (base) transformations of some other sentence structure, or the sum of products of transformations (if its parts are themselves transforms of sentences); if it is not, we say it is the identity transform of itself. If we can suitably mark homonymities, we obtain that every sentence structure is a unique sum of products of transformations. The set of transformations is then a quotient set of the set of sentences, and under the natural mapping of the set of sentences onto the set of transformations, those sentences which are carried into the identity transformation are the kernel of the set of sentences.

[6] But in addition to these, the one-directional transformations are a partial ordering relation.

In English, these kernel sentences have a few simple structures, chiefly *NV* and *NVPN, NVN* and *NVNPN, N is N, N is A, N is PN*. Every sentence, therefore, can be reduced by transformation to one or more of these; and combinations of transformations of these generate every sentence. The complex variety of the unbounded number of English sentences can therefore be described in terms of a small number of kernel structures, a small number of base transformations with their algebra,[7] and a small number of recursive rules for sentence combination and a dictionary.

2. APPLICATION TO INFORMATION RETRIEVAL

The possibility of applying transformations to information retrieval rests on the fact that, by standards of information or meaning which are of interest to science, a sentence carries the same information as does its transform (after allowing for the differences inherent in the particular transformation, such as being a sentence or being a noun phrase). This is not the case, for instance, for belles-lettres or poetry, or for material where association or the use made of the language vehicle is itself a relevant part of the expression or communication. For scientific, factual, and logical material, however, it seems that the relevant information is held constant under transformation, or is varied in a way that depends explicitly on the transformation used. This means that a sentence, or a text, transformed into a sequence of kernels carries approximately the same information as did the original.[8] It is for this reason that a problem like information retrieval, which deals with content, can be treated with formal methods—precisely because they simplify linguistic form while leaving content approximately constant.

The usefulness of transformation to information retrieval is due in part to the fact that scientific report and discussion does not utilize all the structure which is made available by language, such as the differences between transforms of the same kernel, the various verbs and affixes which are roughly synonymous with respect to science, and so on. There is an extra redundancy in language when used in science, and the removal of some of this redundancy, by establishing what distinctions are not utilized, would make possible a more com

[7] There are also various restricted transformations which apply only under particular conditions.

[8] Further work may be desired on the question of information constancy under transformation. For example, one can test experimentally whether kernels are remembered (and forgotten), whereas transformations are largely unnoticed. Or one can consider each transformation and see what changes are involved (what morphemes and orderings are changed) and what meaning attaches to each change.

pact and more lucidly organized storage of the content of scientific writing. This is not the same as a reduction of English sentences to their logical equivalents. The tools of logic are not sufficient for a representation of the statements and problems of science.

Transformations make it possible to store a text as a sequence of kernels. One might ask what advantage there is in this as against merely storing the text in original form. If a searcher asks for anything which interrelates two words a, b, the fact that a, b both occur in the same sentence does not guarantee that there is a relation between them; and a, b may be related while occurring in two neighboring sentences (e.g., if there are certain connectives between the sentences). Hence, either he must be given the whole article, or he will miss some relevant sentences and get some irrelevant ones. In contrast, kernels and their connectors specify the relation among words. Hence it is possible to extract, from a storage containing many articles, precisely those kernels or kernel sequences in which a, b are related (or related in a particular way). It thus seems possible both to store a whole article (transformed), to be called out as such, and also to draw upon it, if desired, for individual kernels separately. Problems attendant upon this are discussed in Section 3.

Once a text has been organized grammatically into a sequence of kernels, a further operation of reduction becomes possible. It is possible to compare corresponding sections of various kernels, and on this basis first to eliminate repetition, and secondly to separate out to some extent the sentences that have different status in respect to retrieval; e.g., sentences which are not worth storing, or sentences which should be stored but not indexed, and perhaps even those which should be used as an abstract of the article. The basis for this is discussed in Section 4.

The methods discussed here do not require the exercise of any knowledge of the subject matter of the article, nor do they require any coding of the relevant concepts of each science.

3. INFORMATIONAL KERNELS

The kernels that are obtained from transformational analysis are determined by the set of transformations that can be found in the given language. It was found that reduction of texts to kernels yielded stretches too small for efficient retrieval. Consider, for example, the sentence:

The optical rotatory power of proteins is very sensitive to the experimental conditions under which it is measured, particularly the wavelength of light which is used.

Transformed down to kernels, this becomes;

connector[9]	*kernel*
	the power is rotatory
	rotation is optical
	“ “ “ of proteins
	“ “ “ very sensitive to the conditions
	“ “ are experimental
wh	“ “ “ measured under “ “
particularly	“ “ “ very sensitive to the wavelength
	“ is of light
wh	“ is used

We would like to obtain larger kernels, preferably of the size and structure that would provide separate kernels for the separate requests of information search. Larger kernels can be obtained simply by omitting some of the transformations, for each omission of a transformation would leave some section or distinction intact. If we can find that certain transformations are responsible for separating out (into different kernels) items that we would like to keep together, we would omit these transformations, and the regular application of the remaining transformations would give us kernels closer to the size and type we want.

This result can be to a large extent obtained by omitting the transformations that separate adjuncts from their centers.[10] As we apply one transformation after another to reduce a sentence (all those which are applicable to that sentence), we reach a stage in which the reduced component sentences have the structure of kernels indicated above (end of Section 1.2) except that the AT and *V* have various adjunct phrases (*A, PN,* etc.) associated with them; e.g., *The rotatory properties of proteins depend on wavelength* (reduced from the example in Section 5). The next applicable transformations here would break this up into:

the properties depend on wavelength
“ “ are of proteins
“ “ “ rotatory

Here *the rotatory properties of proteins* is a noun phrase, with *properties* as center and *of proteins* and *rotatory* as two adjuncts. If for information retrieval purposes we omit these adjunct-extracting transformations, we obtain larger kernels that are closer to retrieval needs.

---

[9] In addition, the sharing of words between kernels (indicated by ditto marks) is also a connector between them.

[10] If for every occurrence of construction *XY* in sentence environment *Z,* we can find also the construction *Y* (but not *X*) alone in environment *Z,* we call *Y* the center of *XY* and *X* the adjunct. Thus in a construction *ANVD* (e.g., *The blue liquid boiled rapidly*.) N and F are the centers, constituting an *NV* kernel *(The liquid boiled),* and *A* is adjunct of *N, D* adjunct of *V*.

However, this overshoots the mark, because some phrases which are grammatically adjuncts may be of independent informational interest. It turns out that many of these are phrases which contain the same words as appear in center positions in neighboring kernels. It is therefore useful to introduce a second type of operation, over and above the transformations. This is to compare the words in every long adjunct with the neighboring center words. The result provides a criterion for further transformation. If the adjunct contains (perhaps two or more) words which are centers in other kernels, that adjunct in transformed into a separate kernel; otherwise it is not.

4. RETRIEVAL STATUS OF KERNELS

The operation of comparison introduced in the preceding paragraph is not hard to carry out once we have kernels, each with its two or three centers and their adjuncts. This same operation, once it is introduced, can be put to wider use in deciding how to treat various kernels.

The various sentences of an article differ in informational status, and even certain sentences which may be of interest to readers of the article may not be requested or useful in retrievals. The distinction is sharper in the case of kernels, because transformations usually separate out what we might consider side remarks, comments about methodology or prior science of the article, and so on, from the kernels that carry the central material of the article. This happens because in many cases the different types of material necessarily occupy different grammatical subsections of the original sentence.

If we now compare the centers of the various kernels, we find that the sections which carry the main material of the article are generally characterized by having certain words occur over and over in various relations to each other. These are the kernels from which the abstract should be selected. The kernels of this type are usually also the ones which are likely to be separately useful in answer to search, so that these kernels should be indexed.

Adjoining these kernels are sections which are adjuncts of them or separate kernels connected to them, and which in many cases contain at most one of the words which are centers of the main kernels. These sections often report conditions, detailed operations, and the like, which apply to the main kernel. This material is not needed in the abstract, and in most cases only its most repeated centers would be wanted in an index for retrieval.

There are also metadiscourse kernels which talk about the main material (e.g., discussing the problems of the investigators). These contain words entirely different from those of the main kernels, except that they often contain one word from a main kernel or a pronoun referring to a main kernel. Such kernels may be omitted from storage except in cases where they are retained as modifiers of a main kernel. In any case they need not be indexed.

Finally, there are in many cases kernels which have a vocabulary of their own, not as repetitively interrelated as the vocabulary of the main kernels. These are often comments about the underlying science of the article, including general methodology; in most cases they need not be stored for retrieval purposes.

This characterization of informational statuses is tentative and rough, but the relevant fact is that properties of the type mentioned in each case can be recognized by means of the comparison operation introduced above. To the extent that there is a correlation between the types of word recurrence and the informational status of kernels, it will be possible to set up the comparison operation in such a way as to make the desired distinctions and thus to determine which kernels are to be stored, which of these are to be indexed, and which abstracted. A great deal of investigation is still required here.

5. STORED TEXTS

The operations indicated above transform a text into a sequence of kernels with standard structure. A convenient way of providing for further operations on the kernels (whether retrieval or further analysis) is to mark off the main internal structures of each kernel, since these will have already been established in the course of applying transformations. Each kernel can be divided into at most five sections:

0.	Connectors (binary operators) to other kernels (e.g., *or, because, however, if…. then*); unary operators on the kernel (e.g., *not, perhaps, surprisingly enough*).
1.	Subject noun phrase: center *N* and its adjuncts.
2.	Verb phrase, including its adjunct *D* and *PN* (preposition plus noun phrase).
3.	Post-verb ("object") *N* or *A* or *PN* phrase: center and its adjuncts.
4.	Adjuncts of the kernel as a whole (usually *PN* or connected sentence).

There are various problems in determining this analysis of each kernel as it is produced by transformations from the original sentence, though in most cases the original sentence structure determines the analysis readily. Nor is this the only way in which kernel structure can be arranged. Of the above sections, 0 and 4 (and perhaps 3) may be empty in a given kernel.

The various kernels in an article, or in a neighborhood within the article, can be compared in order to discover repetitions. If the words of section n of kernel m (written m.n) are identical with the words of section p of kernel q, or if m.n contains a grammatical pronoun of q.p, then in position m.n we merely record the address of q.p instead of repeating the words. If only part of a section is repeated this can be indicated by marking this part (as a subdivisional address) and recording the specific address at the point of repetition. If the repetition is of the corresponding sections 1, 2, or 3 in the preceding kernel (the most frequent situation), we may simply leave the repeating position empty. For example,

kernel address					
m	^{0}	^{1}EF	^{2}G	^{3}H	^{4}J
m + 1	^{0}K	^{1}L	^{2}G	^{3}MF	^{4}

would be recorded:

m	^{0}	^{1}E^{5}F	^{2}G	^{3}H	^{4}J
m + 1	^{0}K	^{1}L	^{2}	^{3}Mm5	^{4}

Articles contain so many repetitions as to make this useful. Similarly, if one kernel contains a whole kernel (or a pronoun of it) within it, we record the address of the repeated kernel in the position which includes it. However, if as often happens the containing kernel is of a metadiscourse type (e.g., *We have found that*...) it would be recorded (in full or in summary form) in the 0 section of the kernel which it contained, or else omitted altogether. As an example, we take a sentence drawn from the same text as the previous examples:

One phase of this research, the dependence of the rotatory properties of

N_1 N_2

proteins on wavelength, is recorded here because it is of special importance

V_3 C $N_1 V_4$

to the problem at hand.

The centers of the noun and verb phrases are marked by N_i, V_j.

The structure of the sentence in terms of these phrases is:

N_1, N_2, V_3 C $N_1 V_4$

Three transformations are applicable here, first separating out the sentence structures, then transforming the constructions within a sentence structure:

1. S_{123} C $S_{14} \leftrightarrow S_{123}$. C S_{14}.
2. N_1, N_2, $V_3 \leftrightarrow N_1$ is N_2.N_1 V_3.

(S_{123} indicates the sentence containing the words marked 1, 2, 3.)

Comparison shows that the words of N_2 recur as centers of other kernels;

therefore N_2 is transformed into a sentence in order to be recorded as a separate kernel:

The dependence of the rotatory properties of proteins on wavelength

V_5n of N_6 P_7 N_8

3. V_5n of $N_6P_7N_8 \leftrightarrow N_6V_5P_7N_8$.

If these three transformations are applied and the section markers are indicated, we obtain:

1. $^{0\ 1}$ One phase of this research ^{2} is ^{3} 2
2. $^{0\ 1}$ The rotatory properties ^{2} of proteins depend on ^{3} wavelength
3. $^{0\ 1}$1.1 ^{2} is recorded here
4. ^{0} because $^{1\ 2}$ is of special importance to the problem at hand

Comparison shows that the words of kernels 1, 3, 4 do not recur in the article; these kernels would not be marked for indexing, and it is most likely that further investigation of their word subclasses would lead to their rejection from storage. On this basis also kernel 1 would be replaced by having the words *one phase of this research is* placed in section 0 of kernel 2.

Various modifications and empirically tested criteria can be added to these operations. For example, if the title consists of N1 plus adjunct *X* (or if the first sentence of a paragraph contains such N1X) then all occurrences of *the* N_1 in the article (or paragraph) can be credited with a repetition of *X,* i.e., *the* $N_1 = N_1X$. If further investigation specifies the situation in which words like *condition, property* have the status of pronouns, i.e., in which they constitute repetitions of something previously stated, then these words could, like pronouns, be replaced by the words they repeat (or rather, by the addresses of those words). If a list of synonyms can be established for various relational verbs, classificatory nouns, and the like, it would be possible to consider each of those words a repetition of any of its synonyms. On this basis, for example, kernel 2 above turns out to be entirely a repetition of other kernels in the same article, and can therefore be omitted. Such a synonym list goes part way toward indicating logical equivalence between sentences, but only in the direction and to the extent that scientific writing actually permits.

When there is a connector between two kernels, each time one of the kernels is retrieved, the connected one will be picked up also (except that a kernel may not have to pick up those which are subordinate to it). If the connector at p refers to a kernel m which is not contiguous, the address of m would be added to the connector at p. The instructions for this arrangement, as for the others mentioned above (such as the section division of kernels) could be obtained as a by-product in the course of carrying out transformations.

In the kernels which are marked for indexing, a suitable criterion can be set up; for example, all words may be indexed except those in a stated list (*the, some,* etc.), or we might index only the center words and the words which are centers of the adjuncts. If someone searches for kernels that connect two words (or their synonyms), the index will yield two long lists of addresses; addresses occurring in both lists give the desired kernels. There are certain strong connectors such that a word occurring in one of the connected kernels may have an important relation to a word in the other (though this is not the case for other connectors or for unconnected kernels). Kernel sequences with such connectors will receive only a single address, so that if a person asks for two words, the index will find the same address under each of them, even if they are in different but strongly connected kernels.

6. MACHINE APPLICATION

The nature of the operations which have been described above is such that they may in principle be performable by machine; in particular, they have been based on the position which words occupy, not on meanings. This is, of course, quite different from discovering the structure of a language, which is also based only on the relative occurrence of physical entities, but is not reducible to such simple formulation. The general question about machine performance of these operations hinges on whether a decision procedure can always be found for the requisite work, and on whether it would be sufficiently short in all cases.

Much of the analysis of language structure is based upon comparison of the positions of a great many words in a great many sentences. This would take too long on a machine. However, the results of this analysis can be built into the machine. For example, there can be a dictionary which gives the class membership of each word, or, where a word is a member of more than one class, determines the class membership on the basis of the class membership of neighboring words. And there can be a list of transformations, each stated in terms of a particular class sequence; if a transformation depends upon particular members of a class this would be indicated under those members in the dictionary. Nevertheless, there may remain various distinctions which are necessary in order to determine the applicable transformation but which cannot be recognized by the machine in terms of the material immediately available to it. The most obvious case is that of homonymities, where no formal division is possible because two different transformational products have the same physical shape.[11]

[11] Where this yields an ambiguity, such that the reader might indeed interpret the form in two ways, we would presumably want to leave the ambiguity unresolved, i.e., not carry out the transformation.

In other cases, an analysis (or a choice between two analyses) is decidable, but on the basis of such a network of subclass relations as a human being can keep in mind but is beyond the storage and time capacities of machine use. In such cases it may be necessary to pre-edit the text, that is to insert at certain points in the text marks which the machine would use in deciding what transformation to use. This editing would require only a practical knowledge of the language, not any special knowledge of the subject matter.

The possibilities then are about as follows. The original text is first read into the machine (perhaps by a print reader, if one can be developed at least for selected fonts in which the major journals are printed). If any pre-editing marks are needed, they would be included in the text; word space, punctuation marks, and paragraph space will of course be noted. The machine then works on a sentence at a time, getting from its dictionary the class and special transformations of each word in the sentence. The representation of the sentence as a sequence of classes (with occasional markers for special transformations) constitutes, with the aid of a stored program for separating out the main phrases, the instructions for carrying out the transformations. (This is not done by a simple left-to-right reading of the marks.) These transformations produce from the original sentence a sequence of tentative kernels, each with its connectors and main grammatical sections marked.

At each paragraph division the machine would institute the comparison operation over the kernels of that paragraph (and perhaps with a check of the main kernels of the preceding paragraph). First, the main words of each long adjunct would be compared with the centers of the various kernels, to see if the adjunct should be transformed into a kernel. The words of each section of each kernel would then be compared in order to replace repetition by kernel-and-section addresses. This would have to be done over stretches longer than one paragraph. Here as elsewhere various simplifications are possible. For example, it might be sufficient to compare the various sections 1 with each other and with the sections 3. There would be few repetitions between these sections and section 0 or 2. Also, the machine might stop this comparison operation if it has gone through say a paragraph plus two sentences (i.e., into the second paragraph) without finding a repetition to the section it is checking. A table of synonyms could be used to extend what is considered a repetition.

Next, the centers of each kernel would be compared to see which kernels have the same centers in different relations (e.g., with different adjuncts), and other characterizing conditions. The results of this comparison would indicate whether a kernel is to be rejected or transformed into a section (chiefly 0) of an adjoining kernel, or stored, and whether it is to be indexed, and perhaps whether it is to be included in the abstract. The kernels would then be stored, and their centers or other indicated words would be marked with their address and sent to the index storage.

The whole of this work—linguistic analysis, formulation of the transformations, characterization of informational status, and machine application—is far from done. The remarks presented above indicate only the results at the time of writing.

7. FURTHER POSSIBILITIES

One advantage in the operations proposed here is that they pave the way for further reduction and organization of scientific texts. Discourse analysis would be a step in this direction; transformations are a preliminary to it, while the comparison operation may be considered the first operation of discourse analysis. If further steps become mechanizable in a reasonable way, it would then be possible to carry out further comparisons and reductions on the stored standardized kernels of the text.

More important, as a by-product of analyzing and storing a great many texts it may be possible to collect experience toward a critique of scientific writing and an indication of useful modifications in language and in discourse structure for scientific writing. Science uses more than logic or mathematics but less than language; and in some respects it uses formulations for which language is not very adequate. However, since scientific communication operates via language (except for mathematical expressions, graphs, and illustrations), detailed investigation of how the language is used gives us some picture of what would be a more useful quasi-linguistic system for science.

Linguistic and Machine Methods for Compiling and Updating the Harvard Automatic Dictionary

A.G.OETTINGER, W.FOUST, V.GIULIANO, K.MAGASSY, and L.MATEJKA

Natural languages, unlike some abstract linguistic systems studied by mathematical logicians or some instruction codes devised by designers of automatic programming systems, have no easily described simple structure. Considerable empirical study is therefore necessary to develop for these languages grammars that are sufficiently precise and comprehensive to serve as the basis for any system of automatic translation. The difficulty of such empirical study is attested to by the almost total absence, after nearly a decade of interest in automatic translation, of any but theoretical discussions of the subject. The empirical work that has been reported is generally the result of laborious manual work and, even where machines have been used, results are based on such limited and carefully selected samples that their significance is doubtful.

Significant research on automatic translation presents such massive data handling problems that, unless automatic machines and associated techniques are used as tools to assist in research from the beginning, chaos is a likely result. Research workers with both adequate qualifications in linguistics and experience in the design and operation of automatic information processing machines are relatively scarce. Careful planning is therefore essential in order to enable the performance of large scale routine tasks by a team of clerical and technical personnel assisted by automatic machines. The necessary automatic machines are presently available in the form of general purpose digital calculators. By the application of present techniques of automatic programming, and by the development of new ones, much of the programming of these cal

A.G.OETTINGER, W.FOUST, V.GIULIANO, K.MAGASSY, and L.MATEJKA Harvard Computation Laboratory, Cambridge, Massachusetts.

culators can systematically be reduced to routine processes to be performed by clerical personnel or even by the machine itself.

While enough is already known about automatic translation and allied problems to warrant paying attention to the development of suitable input, output, and storage devices, the processes of translation are hardly well enough defined yet to justify the construction of a complete specially designed translating system. General purpose machines can provide adequate algorithmic power and sufficient storage capacity with a minimum of capital outlay. Different methods can readily be tested merely by writing different programs, suggesting a method of successive approximations, whereby the results of experimental operation of the system up to a given time can be used to improve the mode of operation for future time. Whenever possible, it is desirable to make program modification itself a part of automatic machine functions. In this fashion optimal design parameters for special equipment can eventually be determined. For a variety of field applications, general purpose machines may well continue to be used even for production, although it is likely that the efficient and economical operation of large translation centers will eventually require the use of specially designed equipment.

With these premises in mind, the work described in this paper has been centered on the formulation and practice of efficient techniques for the initial compilation and periodic up-dating of automatic dictionaries. The first Harvard Automatic Dictionary is intended primarily to provide the following three facilities: *(a)* an immediately useful device for lightening the burden on professional translators, speeding up their work, and improving its accuracy and timeliness; *(b)* a system of automatic word-by-word translation, serving as a linear first approximation to an automatic translation system; *(c)* an experimental tool to facilitate the extensive basic research still necessary to develop methods for faithful smooth translation of technical Russian into English.

While automatic translation from Russian to English is the specific object of our research, automatic dictionaries and the techniques for compiling them lend themselves to more general applications, including translation between other pairs of languages, and certain phases of information organization and retrieval. Automatic abstracting, by techniques such as described by Luhn (1), is one example of such an application. An automatic dictionary in which a record is kept of the frequency of use of each entry can provide an accurate and current standard for eliminating the “noise” caused by words common in any text. The inverse inflection algorithm mentioned in Section 2, extended if desirable to account for derivation as well, can be useful in mapping inflectional variants of a stem, or derivatives of a root, into a single class or canonical form. When the source text is in a foreign language, combining automatic abstract ing with automatic translation of the abstract is an obvious possibility. In searching texts for the occurrence of key words or word combinations, the use of inverse inflection and derivation algorithms will permit specifying keys in a canonical form, with the guarantee that occurrences of inflected or derived variants will also be detected. As the translation algorithms based on the use of automatic dictionaries grow in sophistication, the ties between automatic language translation and the many areas where syntactic analysis and code conversions are necessary will very likely continue to be strengthened.

1. WORD SELECTION

In planning for dictionary compilation, efforts were made to minimize the need for manual intervention, to retain flexibility for experimental purposes, and to provide procedures suitable for periodic up-dating of operating dictionaries, as well as for their initial compilation.

The initial selection of entries for a dictionary can be carried out by two major processes, each having peculiar advantages and disadvantages. The first method may be likened to panning for gold. In this method some number of texts of the type eventually to be translated are scanned, and a glossary of all distinct forms occurring in this sample is compiled. The main advantage of the method is that every form obtained in this way is in current use, and therefore is a suitable candidate for inclusion in a glossary. If scanning is to be performed automatically, the texts must necessarily be transcribed onto some automatically readable medium, hence made available for other purposes, including the eventual testing of translation methods. The major disadvantage of the procedure is that it becomes progressively less and less productive. While nearly all the first few words of the first text scanned are likely to be nuggets, as the work progresses more and more gravel must be handled before another nugget is found. Relatively few word forms account for the vast majority of form occurrences in any text; these forms are found over and over again and must be rejected over and over again. Second, in pure panning, only those particular inflected forms of any word that actually occur in the sample under study will be entered in the glossary. Until all the forms constituting the paradigm of a word have occurred in some text, a complete characterization of this word is not available. This precludes the possibility of early systematic treatment of inflectional processes, other than by handling each inflected form as a unique entity.

The second approach may be called the “fish net” method. Available dictionaries are dragged for words useful according to some reasonable criterion. Anything caught in the net is retained. The chief advantage of this procedure is that a large selection of useful words is obtained very rapidly. The major disadvantage is that the resulting dictionary is only as good as the criterion of selection or as the dictionaries from which it was selected. Moreover, words without current utility may well be caught in the net.

Neither method is guaranteed to yield a vocabulary precisely suitable for the first texts to which a dictionary is applied, but either lends itself to the addition of new words or new forms whenever these appear in a text. How rapidly the ratio of text words not in the dictionary to those in it will diminish to a satisfactory level is still open to conjecture. For that matter, so is a satisfactory definition of this level.

The procedure described here is a combination of the two methods. An initial set of nearly ten thousand words was selected, in part, from a general dictionary (2) to obtain words of common currency and, in part, from a specialized electronics dictionary (3) to obtain as complete as possible a coverage of technical terms in this area. In cases of doubt about the utility of words, they were usually included. It seems more efficient to carry a few doubtful words through routine compilation procedures and to provide for their automatic removal (based on a criterion of frequency of use), once operating experience has accumulated, than to spend valuable personnel time on intricate and inconclusive selection procedures. Once the dictionary is functioning, all new words encountered in texts submitted for translation, but not in the dictionary, will be printed as a by-product of dictionary operation. These new words, from text sources, will eventually replace the initial stock as the raw material for an up-dating procedure almost identical to the compilation process. For up-dating, therefore, the fish net process is replaced by a modified gold panning technique, where the gravel is the major object of processing, and the gold a valuable, but easily obtained by-product; prior to further processing, each new form found in a text is reduced to the form normally listed in dictionaries. The accumulation of frequency data, the analysis of contexts and other analytic procedures can be carried out on texts already recorded in automatically readable form because of their interest as objects of translation.

2. THE FORM OF DICTIONARY ENTRIES

In ordinary dictionaries, the paradigm of a word is conventionally represented by a standard or “canonical” form, e.g., the nominative singular for nouns, the nominative masculine singular for adjectives, and the infinitive for verbs. This lexicographic device presupposes on the part of a person using the dictionary the ability to perform the grammatical analysis necessary to reduce an inflected form of a word found in a text to the canonical form listed in the dictionary. As conventionally practiced, this grammatical analysis requires a fairly thorough acquaintance with the inflection system of the language in question, as well as a certain amount of imagination.

An automatic dictionary may provide for the treatment of inflected forms either *(a)* by providing a distinct entry for each distinct inflected form or *(b)* by providing only a single canonical form as an entry for each word, together with an algorithm for transforming other forms to the canonical one. A variety of compromises between these two extremes is also possible. Whichever type of entry is used, if words are selected from existing dictionaries by the fishing technique, an algorithm capable of generating all distinct inflected forms of every word given in the conventional canonical form is essential. The need is obvious if a distinct entry is to be made for each distinct inflected form. Under circumstances described further on, the generation of all distinct inflected forms is useful even if only the canonical forms are to be used as dictionary entries. The pure gold panning procedure by-passes these difficulties, but at the price of commitment to the system of distinct entries for distinct inflected forms.

If experimental work is to be carried out on existing machines, facilities sufficiently ample and economical to store a large dictionary are presently available only in the form of magnetic tapes or punched cards. Because using distinct inflected form entries would increase storage requirements and search time by an order of magnitude, this method will be practicable only when large capacity internal memory devices with sufficiently low cost and access time become available (4, 5). Certainly for the present, and probably for the future, the use of canonical form entries seems indicated.

The canonical form chosen for our purposes is best described as a "stem," because its definition is very close to that of stems as commonly defined by linguists. The precise definition of stems and a description of the algorithm used for reducing arbitrary members of a paradigm to this canonical form have been given earlier (6). The reduction algorithm or "inverse inflection algorithm" cuts forms into two parts, one, the stem, the other, an ending usually identical with or at least very similar to the usual Russian inflectional desinences. When the dictionary is in operation, each form occurring in a text is split by the inverse inflection algorithm into a stem and an ending. The stem is used as the key for search in the dictionary, and the ending is retained for eventual use in syntactic analysis. To insure that the stems produced by splitting text forms will be identical with stems entered in the dictionary, the latter are obtained by applying the inverse inflection algorithm to all members of the paradigm of any word about to be entered in the dictionary.

3. THE GENERATION OF PARADIGMS

Generating all inflected forms belonging to the paradigms of ten thousand canonical forms obtained from standard dictionaries is a task not to be undertaken lightly. As a purely manual operation, this task would be staggering, and the probability of errors extremely high. The standard Russian noun paradigm has twelve forms, the adjective paradigm twenty-four or twenty-eight forms, and verbs have well over one hundred forms, if participles are counted. Because not all members of every paradigm are distinct (e.g., the nominative singular form of some nouns is identical with the accusative singular form), it is sufficient to generate a "condensed paradigm," containing only distinct inflected forms. However, even the number of distinct inflected forms is enormous.

In Russian, inflected forms are characterized by their desinences, so that the specification of a paradigm is tantamount to specifying a type of arrangement of inflectional desinences. While the morphological differentiation of word stems is great, that of inflectional desinences is relatively small. The inflection pattern of one word can therefore be used to obtain the inflected forms of all other words characterized by the same nature and disposition of inflectional desinences. This suggests the possibility of developing an algorithm for automatic direct inflection, which will treat alike all words identified as having identical or sufficiently similar inflection patterns.

A new system of classification was therefore developed (7) such that, once a word has been identified as belonging to a class whose members are inflected in a certain way, it can be inflected automatically. Classification systems given in existing grammars were found to be both incomplete and often incompatible with our requirements. Our system of classification is based on the assumption that the identification of the inflectional pattern of a word must, for the time being at least, remain a manual function, while the actual generation of distinct inflected forms can be an almost completely automatic process. Therefore, ease and accuracy of identification must be promoted by using any readily obtainable data meaningful to a person, while the generation must be based strictly on explicit orthographic data recognizable by a machine.

Some of the criteria used in defining classes may be illustrated with reference to Fig. 1. The first column of Fig. 1 describes the full paradigms of the Russian words дама, комната, and граница. Condensed paradigms, where only distinct inflected forms appear, are illustrated in the second column. It will be noted that дама and комната have been included in the same class, even though they differ in the accusative plural, because their condensed paradigms can be generated by identical procedures. The justification for including **граница** in this class is twofold: first, differentiating **граница** from **комната** is inefficient at this stage and is accomplished more readily at a stage described in Section 4. Second, the only price to be paid for simplifying classification in this way is the addition of a spurious inflected form to the paradigm of each member of this class. This addition is harmless. If a dictionary of distinct inflected forms is being compiled, the spurious forms will never be consulted and therefore will eventually be eliminated by virtue of their zero frequency of use. In a dictionary of canonical form stem entries, a spurious form usually leaves no traces, since it leads to a stem identical to those obtained from the other distinct inflected forms. This procedure is very much like one used quite frequently in approximating mathematical functions over a given range: any convenient function may be used that suitably approximates the desired function in the specified range; its behavior outside of this range is of no consequence. Other expedients of this kind are also proving their worth in terms

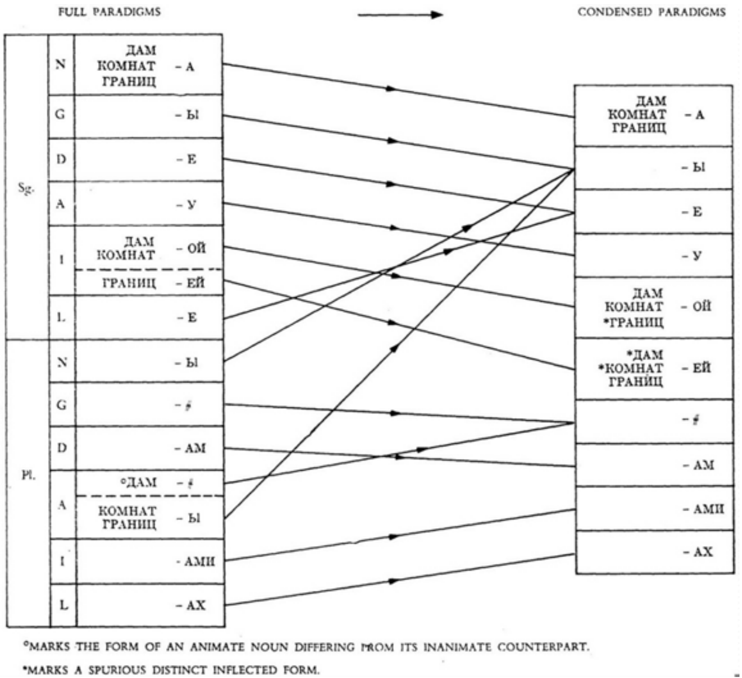


FIGURE 1. Condensation of paradigms (examples drawn from class N4).

of greater systematization than would be possible without them. As those experienced in automatic data processing well know, the handling of a single exceptional case often proves more costly and more time-consuming than the routine handling of a few extra elements.

The definitions of inflectional classes and the rules for forming distinct inflected forms of words belonging to these classes are illustrated in Fig. 2. The characteristics identifying a class are given under (I) and the process of inflection is described under (II). It will be noted that all structural operations to be performed in the process of automatic inflection, or the phenomena bearing upon them, are described strictly in terms of orthography. For example, in the class N4, the set of distinct inflected forms is described as consisting of the standard dictionary canonical form, plus several other forms generated by adding specified endings to a generating stem. The rule of formation for the generating stem itself is given in such terms as "canonical form minus last letter." The generating stems defined in the formation rules for distinct inflected forms are not necessarily identical with the stems that will be used as entries for the dictionary. The latter are obtained by applying the inverse inflection algorithm

N4. дама, лампа, игла, служба

I. A class ending in -a preceded by any consonant except:
- (i) г, ж, к, х, ч, ш, щ;
- (ii) ц, whenever preceded by another consonant;
- (iii) the majority of cases when the consonant is л, н, } preceded by another consonant
- (iv) some cases when the consonant is м, р, } preceded by another consonant
- (v) a few cases when the consonant is б, } preceded by another consonant

II. Formation rules for distinct inflected forms:
- (i) Generating stem = canonical form[a] – last letter;
- (ii) Distinct inflected forms:

(a) canonical form		(f) generating stem	+ ей
(b) generating stem	+ ы	(g)	+ #
(c)	+ е	(h)	+ ам
(d)	+ у	(i)	+ ами
(e)	+ ой	(j)	+ ах

N4.31 выплавка, кишка́

I. A "reappearing o" class embracing the nouns ending in –a preceded by:
- (i) any consonant (except й, ж, ш, ч, ц), not followed by ь, + к;
- (ii) ж, ш, ч, + к, whenever the stress falls upon the ultima of the word;

II. Formation rules for distinct inflected forms:
- (i) Generating stem = canonical form – last letter, but an o must be inserted between the penult and the ultima of the generating stem in (g);
- (ii) Distinct inflected forms as in N4, II (ii), except -и for -ы in (b).

[a] The canonical form here is that used in standard dictionaries, not the "stem" canonical form used in the Automatic Dictionary.

FIGURE 2. Rules for class identification and inflection.

to the distinct inflected forms, and the resulting canonical stems are not always identical with the generating stems. For example, the past passive participial forms of some verbs are constructed most readily by adding –енный to a generating stem. The canonical stem of the corresponding paradigm includes the letters –енн.

Because the assignment of words to classes is intended, for the present, to be a manual task, any classification criterion that is easily recognized by persons can be used. For example, stress distinctions, which cannot be used in defining formation rules, serve as a means of class identification. In addition, significant examples, lists of exceptions, and so forth, have been given wherever possible.

Our system also embraces a large number of words whose formation is "irregular." For example, the class N4.31 comprises words in which the vowel "o" is introduced in one inflected form. The formation rules for this class (Fig. 2) therefore specify that an "o" must be inserted between the last and next to the last letters of the generating stem before constructing the form (g). The class N4.31 is further distinguished from the class N4 by the use of the letter и for the letter ы in the inflected form (b). Whenever the formation rules for one class deviate only slightly from those for another class, they have been stated as exceptions to the rules for this other class. Considerable economies in programming inflection are achieved as a result.

Our system of classification comprises eight classes of adjectives, thirty-eight classes of nouns, and forty-six classes of verbs. Indeclinable words are assigned to a special class. The system of classification is sufficiently comprehensive to include all but a few unproductive classes with highly atypical paradigms; it is completed by the definition of a class, labeled Z99.99, to which all words not falling into any of the other classes are assigned. These words are eventually inflected by hand. Of a total of seventy-six hundred words classified to date, all but twenty-four were assigned to genuine classes. The distribution of these words among the major groupings is indicated in Fig. 3.

Total number of:		
Adjectives	2477	32.59
Nouns	3972	52.26
Invariables	74	0.97
Unclassified nouns and adjectives (Z99.99)	13	0.17
Verbs	1053	13.86
Unclassified verbs (Z99.99)	11	0.14
Total	7600	99.99

FIGURE 3. Distribution of words among major groups.

The class definitions as given in Fig. 2 are not in a form that readily lends itself to rapid recognition. To make classifying as easy as possible, the rules were expressed in a tabular form (3) illustrated in Fig. 4, where the complete classification table for adjectives is displayed. The vertical lines in Fig. 4 divide endings on the right from significant terminal stem letters on the left. The special symbols C, V, and C_bC are interpreted as follows: C denotes any consonant not specified earlier in the group between horizontal lines, and V the same for vowels. The combination C_bC signifies any consonant, whether followed or not by the soft sign, and not specified earlier within the group. For example, an adjective ending in –ий preceded by к may be assigned to any of the classes A6, A7, or A8, depending on the letter preceding к. If the letter preceding к is one of the three indicated next to A7, the adjective is assigned to this class, if the letter preceding к is any consonant not in the list belonging to A7, the adjective is assigned to the class A6, and finally, if the letter preceding к is any vowel, the adjective is assigned to the class A8. Vowel changes are marked by the sign >. On the left side of the sign is the vowel before the change, and on the right side the vowel after the change. For example, in the third group in Fig. 4 adjectives ending in –ий preceded by π or p are normally assigned to the class A3, unless the vowel e is inserted as the second letter from the end in the masculine predicative form of the adjective. Through the use of these charts, assigning to classes becomes a routine task which can be done by a person with relatively little knowledge of Russian, although occasionally dictionaries must be consulted in the process. Experience has shown that the amount of dictionary consultation at this stage is negligible. With this one exception, all inevitable dictionary consultation is concentrated at a single stage of the process, namely, when English correspondents are assigned and grammatical codes are added to the stems. Assigning words to classes has been successfully done at an average rate of approximately one thousand words per day per person.

4. THE PREPARATION OF STEM ENTRIES

Because our dictionary is intended for operation on the Univac I computer at the Harvard Computation Laboratory, some details of the preparation of stem entries and of other phases of compiling apply directly only to this machine. However, very similar procedures can be readily used with other types of contemporary large-scale computers.

Words selected from the two dictionaries mentioned earlier were originally transcribed onto file cards. An inflectional class marker, assigned in the manner outlined in Section 3, was then written on each card. The card file is used only in the initial compilation process, since new words found in texts as a by-product of dictionary operation will be made available automatically in a

ADJECTIVES

				1
	ш	ий	A4	
	ж	ий		
	ч	ий		
	щ	ий		
	н	ий	A5	
	г	ий	A8	
	х	ий		
	жк	ий	A7	
C →	йк	ий		
	ьк	ий		
C →	к	ий	A6	
V →	к	ий	A8	
				2
	ж	ой	A8	
	к	ой		
	ш	ой		
	г	ой		
	х	ой		
(C_bC) →	н	ой	A2	
V →	н	ой	A3	
C →		ой	A3	
				3
	нн	ый	A1	
(C_bC) →	н	ый	A2	
V →	н	ый	A3	
	л	ый	A3 but A2 if masculine predicate: $\text{ø} > e_2$	
	р	ый		
C →		ый	A3	

FIGURE 4. Classification table for adjectives.

printed format suitable as raw material for the periodic dictionary up-dating.

Classified words are next recorded on magnetic tapes by means of a standard Unityper. Because this input device is designed to handle normally only numerals, the Roman alphabet, and a few special characters, some adaptation was necessary to use it for typing Cyrillic material. The set of characters used for this purpose is shown in column 1 of Fig. 5. It is a fairly simple matter to place over the normal keys of any typewriter special keytops engraved with any desired alphabet. We chose to arrange the keyboard in one of the standard Cyrillic layouts, to ease the work of typists already familiar with it. The correspondence so established between Cyrillic and machine characters is described by columns 1 and 2 of Fig. 5. Although this correspondence preserves a familiar layout, it does not preserve normal Cyrillic alphabetic order, and alphabetization of words in the code of column 2 is impossible. Magnetic tapes obtained from the typewriter are therefore used as input to a code conversion run in which the typewriter code of column 2 is converted into the ranked code given in column 3. The correspondence between Cyrillic characters and machine characters becomes that given between columns 1 and 3. The ranked code of column 3 is used throughout compilation and up-dating ell aas ws throughout the dictionary look-up operations. Russian material re

UNIVAC Codes for Cyrillic Alphabet

1	2	3	4	1	2	3	4	1	2	3	4
0	0	,	0	З	O	8	Z	Ш	U	R	SH
1	1	&	1	И	B	9	I	Щ	I	+	SHCH
2	2	r	2	Й	Q	;	J	Ъ	r	S	#
3	3	A	3	К	E	B	K	Ы	S	T	Y
4	4	F	4	Л	K	C	L	Ь	M	U	'
5	5	¢	5	М	V	D	M	Э	&	V	EH
6	6	@	6	Н	T	E	N	Ю	.	X	JU
7	7	t	7	О	J	G	O	Я	Σ	Y	JA
8	8	/	8	П	G	H	P	Δ	Δ	Δ,0	Δ
9	9	J	9	Р	H	I	R	(	(	(	(
А	F	1	A	С	C	)	S	#	#	#	)
Б	,	2	B	Т	N	K	T	"	"	"	;
В	D	3	V	У	W	L	U	$	$	$	:
Г	Y	4	G	Ф	A	M	F	*	*	*	*
Д	L	5	D	Х	P	N	X	.	+	.	.
Е	R	6	E	Ц	–	P	TS	,	|	,	,
Ж	;	7	ZH	Ч	X	Q	CH	%	%	%	–

1: Cyrillic Available on Keyboard
2: Typewriter Code
3: Ranked Code
4: Transliteration Code

FIGURE 5.

corded in the ranked code obviously cannot be easily read, so that material which must be read quickly is subjected to still another code conversion run in which the character strings given in column 4 of Fig. 5 are substituted for their correspondents in column 3. Eventually, the conversion from the code of column 1 to that of column 3 will be made simultaneous with typing.

Two lists are prepared from the tape recorded on the typewriter. The first, illustrated in Fig, 6, presents the words in the order in which they were typed

ESLI	I01.00	004-0059 004		006050
ESHCHE	I01.00	004-0060 003		006059
ZHE	I01.00	004-0061 002		007000
ZHELANIE	N10.00	004-0062 007		007005
ZHIRNYJ	A02.00	004-0063 006		007010
ZHURNAL	N01.00	004-0064 006		007015
ZA	I01.00	004-0065 002		007020
ZAVISET'	V06.20	004-0066 008		007025
ZAVYSHENNYJ	A01.00	004-0067 010		007030
ZADACHA	N08.10	004-0068 006		007035
ZAKLJUCHENIE	N10.00	004-0069 010		007040
ZAKON	N01.00	004-0070 005		007045
ZAMETIT'	V04.01	004-0071 008		007050
ZAPIRAT'	V01.00	004-0072 008		007055
ZAPUSK	N01.10	004-0073 006		008000
ZARJAD	N01.00	004-0074 005		008005
ZATUXANIE	N10.00	004-0075 009		008010
ZVUKOVOJ	A03.00	004-0076 008		008015
ZDES'	I01.00	004-0077 005		008020
ZNACHENIE	N10.00	004-0078 008		008025
ZNACHITEL'NY J	A02.00	004-0079 012		008030
ZNACHITEL'NO	A02.00	004-0080 011		008035
I	I01.00	004-0081 001		008040
IDENTICHNYJ	A02.00	004-0082 010		008045
IDTI	Z99.99	004-0083 004		008050
IZ	I01.00	004-0084 002		008055
IZO	I01.00	004-0085 003		009000
IZVNE	I01.00	004-0086 005		009005
IZMENENIE	N10.00	004-0087 009		009010
IZMENIT'	V04.00	004-0088 008		009015
IZMENIT'SJA	V04.00	004-0089 010		009020
IZMERIT'	V04.01	004-0090 008		009025
IZOLIROVANNY J	A01.00	004-0091 101		009030
IZOLJATSIJA	N07.00	004-0092 008		009035
IZUCHAT'	V01.00	004-0093 007		009040
IZUCHENIE	N10.00	004-0094 008		009045
ILI	I01.00	004-0095 003		009050
IMENNO	I01.00	004-0096 006		009055
IMET'	V01.00	004-0097 005		010000
IMPUL'S	N01.00	004-0098 007		010005
IMPUL'SNYJ	A02.00	004-0099 010		010010
INSTITUT	N01.00	004-0100 008		010015
ISKAZHENIE	N10.00	004-0101 009		010020
ISKLJUCHAT'	V01.00	004-0102 009		010025
ISKUSSTVENNY J	A01.00	004-0103 101		010030
ISPOL'ZOVAT'	V03.00	004-0104 012		010035
ISPYTANIE	N10.00	004-0105 009		010040
ISTOCHNIK	N01.10	004-0106 008		010045
ISXODNYJ	A02.00	004-0107 008		010050
ITTI	V16.00	004-0108 004		010055
IDITI	V16.00	004-0109 005	ПОЙТИ	011000
IX	Z99.99	004-0110 002		011005
K	I01.00	004-0111 001		011010
KO	I01.00	004-0112 002		011015
KAZHDYJ	A03.00	004-0113 006		011020
KAK	I01.00	004-0114 003		011025
KAKOJ	A08.00	004-0115 005		011030
KASKAD	N01.00	004-0116 006		011035
KATOD	N01.00	004-0117 005		011040
KOTODNYJ	A02.00	004-0118 008	КАТОДНЫЙ	011045
KACHESTVO	N08.00	004-0119 008		011050
KVADRAT	N01.00	004-0120 007		011055
KVADRATICHNY J	A02.00	004-0121 012		012000
KVADRATNYJ	A02.00	004-0122 010		012005
KOGDA	I01.00	004-0123 005		012010
KOLEBANIE	N10.00	004-0124 009		012015

FIGURE 6. Alphabetized word list.

and is used for catching errors made in typing the words; it will eventually be replaced by hard copy made by the typewriter itself. The other, shown in Fig. 7, is alphabetized by the last letters of words rather than by the first as usual, and justified to the right to bring all words with like endings together. It is used to check the classification of words. Because the criteria for classification are based largely on the configuration of the last letters of each word, words with the same class marker tend to be brought together on this list.

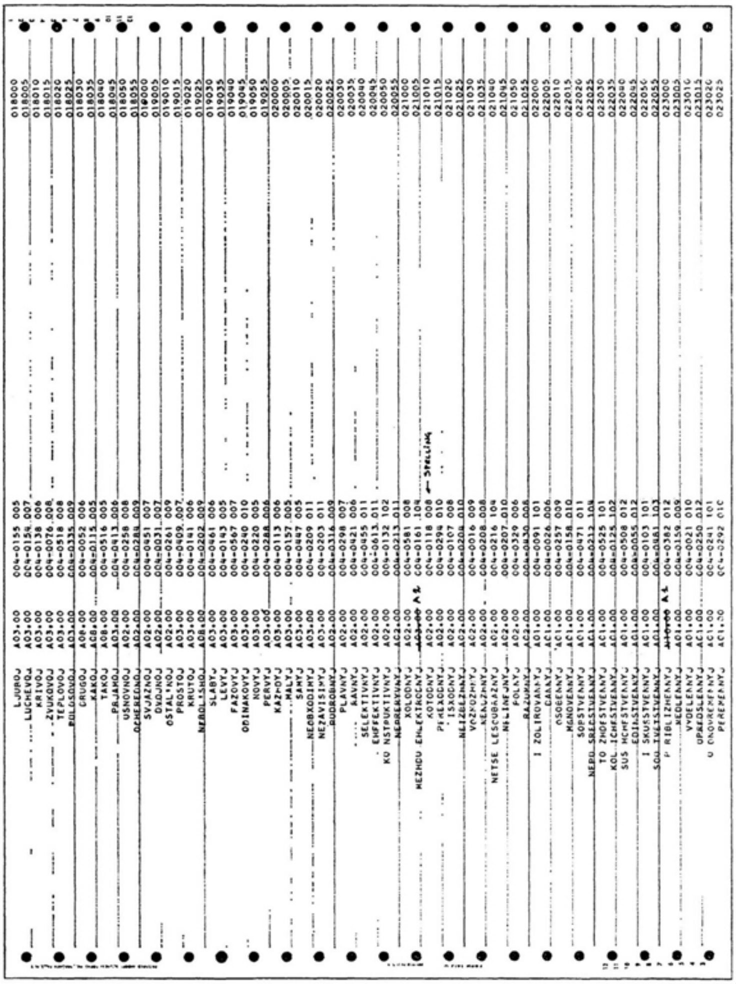


FIGURE 7. End alphabetized word list.

Checking class markers is facilitated by the presence of runs of identical class markers, in which the occurrence of an odd marker shows up clearly. Errors detected on these lists are not corrected on tape. It is easier to delete the affected items prior to assigning correspondents to them, and then to process them again in routine fashion with another batch. Therefore, items found in error are simply marked for deletion at the later stage.

Throughout compilation (8), Russian words represented in the ranked code are imbedded in “items” consisting of five machine words of twelve characters each. The first three machine words are used to store the Russian word, allowing a maximum of thirty-six letters. The first six character positions of the fourth word of each item are reserved for the inflectional class marker as shown in Figs.6 and 7. Eight positions in the fifth word are used for an identification number which designates the batch in which the word was originally processed, and the serial number of the word within this batch. This identification number accompanies all forms of a word throughout compilation, to facilitate the identification of Russian words represented in the ranked code, and the tracing of errors. Of the last three digits of the fifth machine word, the two low order ones specify the character position within a machine word at which the last letter of the Russian word occurs, while the high order digit specifies the machine word (0, 1, or 2) in which the last letter of the Russian word occurs. These numbers are computed during the initial code conversion, and are used to control shifting and extracting operations when the Russian word undergoes later transformations.

After the lists of Figs. 6 and 7 have been checked for errors, the Russian word tapes are ready for automatic inflection (9). This process is illustrated in Fig. 8, with the verb **писать** as an example. This verb is shown with its inflectional class marker V5, assigned as described in Section 3. The forms generated after one inflection run are indicated in column α of Fig. 8. It will be noted that adjectival inflectional class markers are given next to the participial standard canonical forms. These forms are generated by the inflector routine, together with their class markers. Therefore, the inflection of the participles is entirely automatic, without need for further manual classification.

The inverse inflection algorithm is then automatically applied to the set of distinct inflected forms, splitting each into an ending and a potential stem canonical form. The resulting set of stems is condensed to yield a list where only one representative of each distinct type of stem is retained. As in the example of Fig. 1, and for the same reasons, some artificial forms are generated which do not properly belong to the paradigm of the verb **писать**. These forms are marked on the diagram by the letter I. In addition, several so-called academic forms are generated. These are distinguished from the artificial forms in that they are morphologically acceptable members of the paradigm of **писать** but are very unlikely to occur in actual texts. Stems obtained from artificial or academic forms can be identified and deleted, if desired, at the time that English correspondents are assigned. In case of doubt they are retained, to be deleted when their zero frequency of use after a long period of operation automatically indicates that they should be.

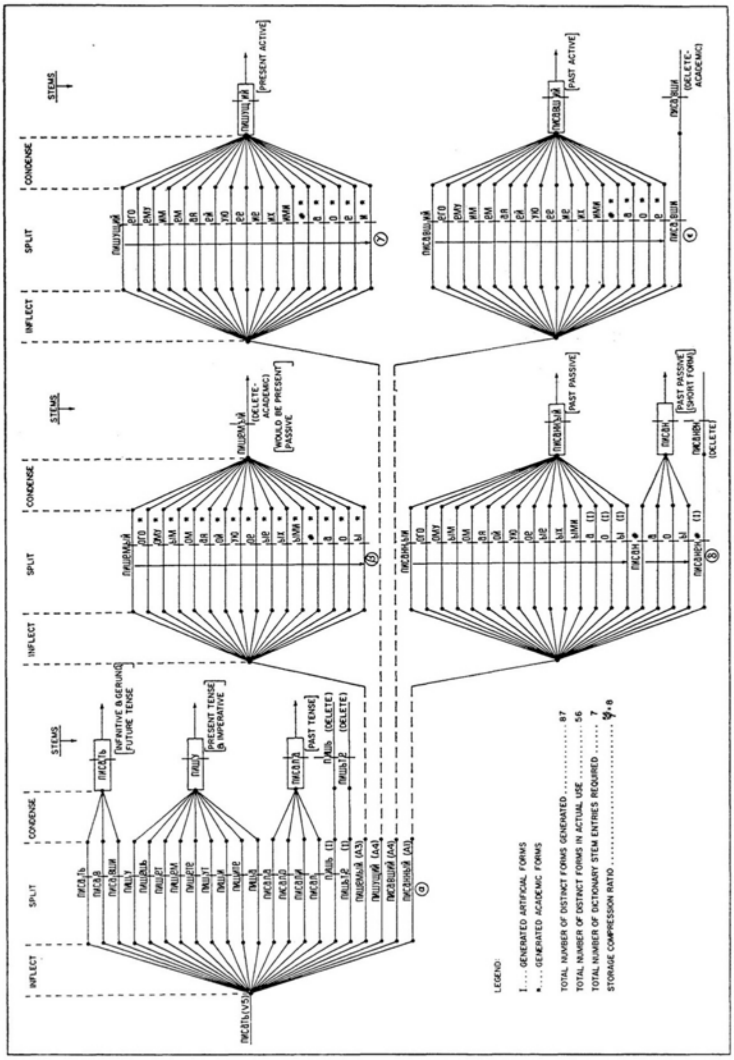


FIGURE 8. The generation of paradigms and canonical stems.

In practice, automatic inflection is separated into three major runs. Words marked with nominal and verbal class markers are inflected before those with adjectival class markers. This means that, on the last run, those adjectival forms generated together with their class markers as a result of verb inflection can be inflected with the other adjectives. A list of inflected forms obtained from the inflector runs is shown in Fig. 9. A similar list, showing forms after splitting of their endings, is given in Fig. 10. Adjectival forms generated as a result of verb inflection are distinguished by the presence of one of the letters A, B, C,

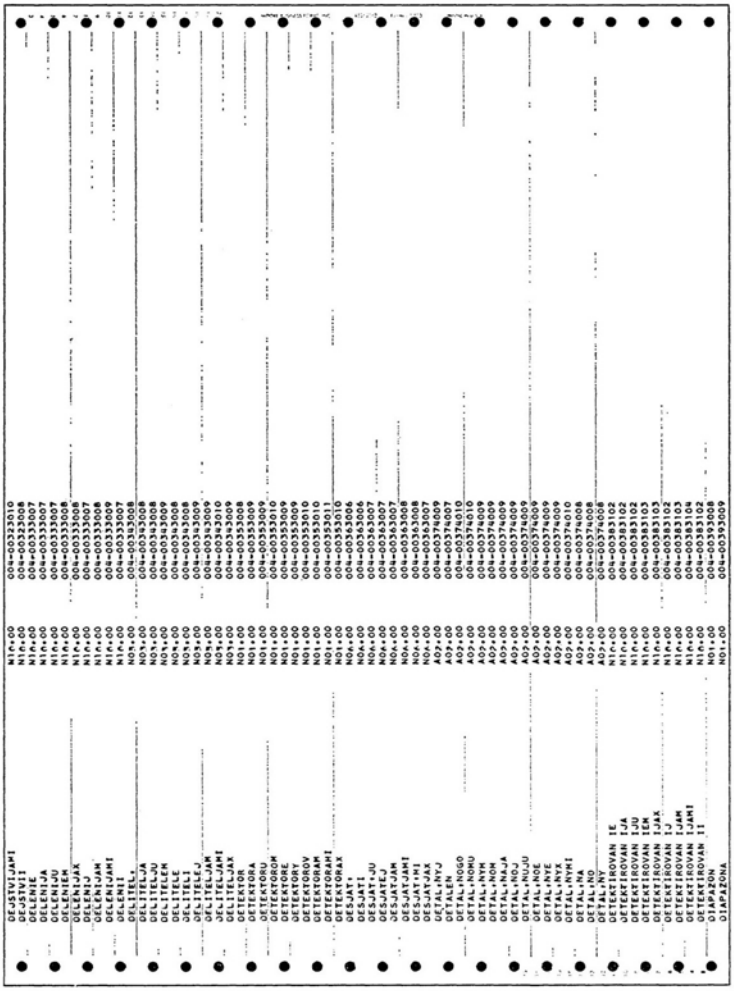


FIGURE 9. Inflected forms.

or D in their serial number, in the position where a hyphen occurs in all other words. On the magnetic tapes used for further processing, only a stem is present in the first three machine words. The split ending is stored in the last five character positions of the fourth machine word of the item, where it may be seen in ranked code. Of these five characters, three are reserved for normal endings, the last two are spaces except for verbs ending in −ся or −сь. For ease in reading transliterated lists of split forms, the endings stored in these

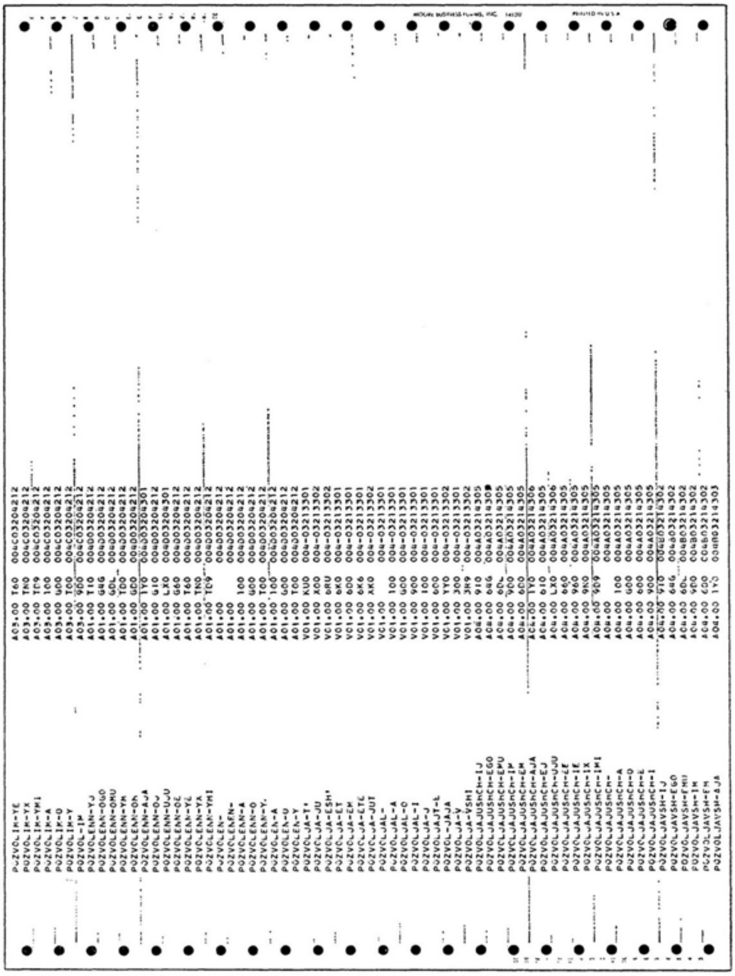


FIGURE 10. Split inflected forms.

positions are brought back into the first three machine words during the transliteration run. The hyphen is inserted to mark the position of the split.

The list of distinct stems obtained by automatically condensing a list of the type illustrated in Fig. 10 is printed with the layout shown in Fig. 11. This layout is designed to guide the manual inscription of English correspondents and of grammatical coding associated with the stem canonical forms. The paper is ruled to show clearly the divisions between machine words; exactly

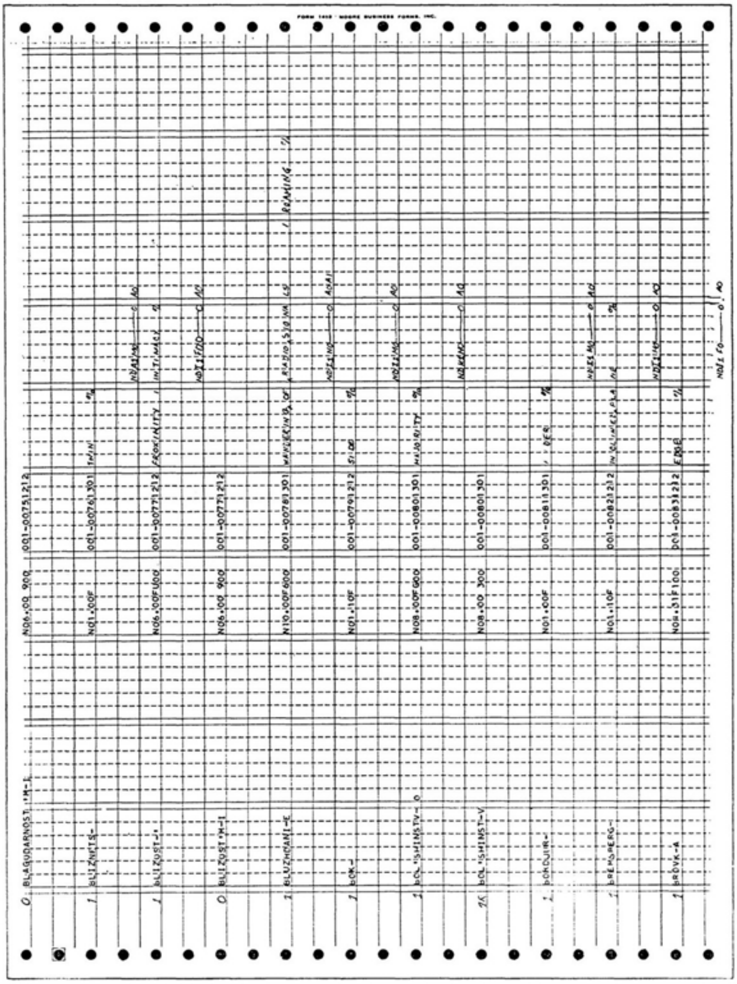


FIGURE 11. Dictionary work sheet.

two characters must be written in the spaces defined by the vertical dashed lines. Heavy black horizontal lines delimit the space allowed for an entry. Dictionary entries are items consisting of thirty machine words, of which the first five make up a standard Russian item. The last twenty-five machine words are devoted to correspondents and to grammatical information. The English correspondents are written immediately after the Russian item. The last four of the twenty-five words are reserved for coded information. Distinct correspondents are numbered, the last correspondent being marked by the use of a percent sign in place of a numeral.

An effort is made to rank the correspondents in the order of their likely frequency of use. Initially, this ranking is necessarily somewhat arbitrary. It is expected that, as a major by-product of automatic dictionary operation, the sets of correspondents and their ordering will gradually be adjusted in accordance with the experience of the technical experts who are the ultimate users of translations and the best judges of their value.

In the first of the four words reserved for coded information, the significance of a character depends on its position within the word. This word is therefore called the "organized word." In the other three words, the significance of characters is independent of their position within the set of three words. These words are therefore called "semiorganized words." Because the inflectional class markers have chiefly formal significance, one of the major functions of the coded information in the organized word is to identify the functional role of the entry. This may be illustrated by the following two examples: a word declined like an adjective may function as a noun and will be coded as such in the organized word; the set of words classified as formally indeclinable includes, along with prepositions and conjunctions, some words functioning as nouns, and these are distinguished by an appropriate notation. Functionally distinct paradigms lumped into one inflectional class to simplify classification and automatic inflection are also distinguished by means of a notation in the organized word. For example, the distinction between animate and inanimate nouns is of no consequence so far as the generation of distinct inflected forms is concerned, but it is vital to the interpretation of the functional significance of endings. That distinction is therefore made by means of a symbol in the organized word. Such information as the names of dictionaries or texts consulted in preparing the entry is mentioned in the semiorganized word. In addition, important grammatical data, whose range of application is insufficient to justify their inclusion in the organized word, are introduced in the semiorganized words.

Although every effort is made to provide for coded information of wide scope, it is clearly not possible to foresee all contingencies likely to be met in experimental work. Therefore, the space between the percent sign marking the end of the last English correspondent and the organized word is left free for the insertion of ad lib English prose comments. In this way, information of significance whose need was not foreseen in planning the layout of the organized and semiorganized words, or likely to apply to too few entries to warrant inclusion in these words, may be recorded and retrieved automatically at a later date. The systematic inclusion of such information into the organized words is possible whenever large enough classes of similar comments are found.

After the correspondents and coded information have been written as shown in Fig. 11, the handwritten material is transcribed onto a magnetic tape. This tape is then merged with that containing the Russian stems, and the complete dictionary entries are recorded on a new tape. The markings in the left-hand margin of Fig. 11 are used to control in part the process of dictionary assembly. A zero in the left-hand margin indicates that the stem is to be deleted and not to appear in the final dictionary. This is the means for deleting spurious or academic forms, or errors detected earlier but not then corrected.

The paradigms of nouns and adjectives where vowel insertion occurs, and of many verbs, may be represented in the list of Fig. 11 by more than one stem. This problem of multiple stems has been discussed in detail elsewhere (6). Whenever two or more stems belong to the same paradigm, they are identified by the same serial number. The first stem in such a set is always that split from the standard dictionary canonical form, and is marked by the letter F in the seventh column of the fourth machine word of the Russian item. Unless it is to be deleted, this stem is usually given a left-hand margin marker 1. If other stems with the same serial number require precisely the same English correspondents and coded information, repeated writing of this information is not necessary. It is sufficient to write the symbol 1R in the left-hand margin for such entries. Where one or more of a set of stems with the same serial number require a set of correspondents distinct from that assigned to the F form, the margin markers 2, 2R, 3, 3R, etc., may be used.

In transcribing the English and coded material, each entry is identified simply by its left-hand margin marker and the serial number. The whole Russian entry need not be copied.

Once the correspondents have been transcribed onto magnetic tapes, they are automatically assembled with their stems. A section of assembled dictionary is shown in Fig. 12. It should be pointed out again in this connection that Russian word forms in dictionary entries are stored in the ranked code on magnetic tape and are represented by stems only. Endings are introduced during the process of transliteration only to facilitate human recognition of the printed entries.

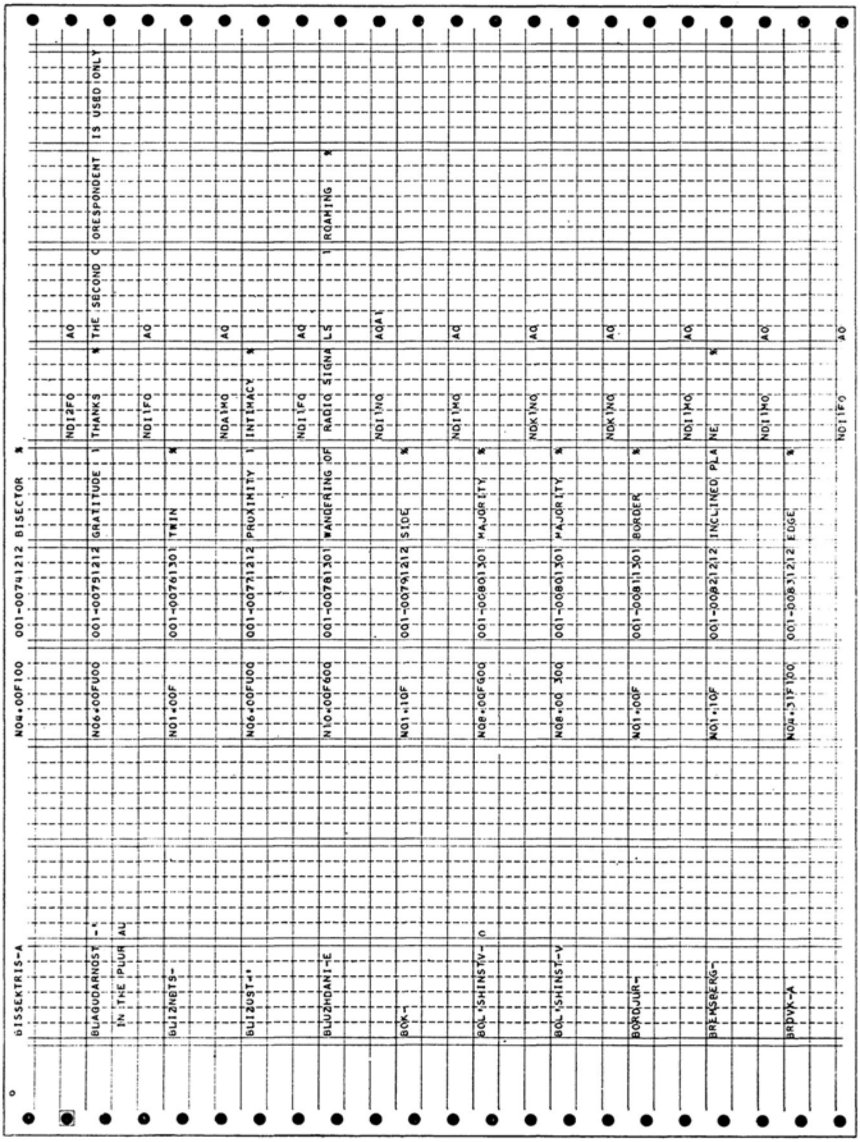


FIGURE 12. Assembled dictionary.

ACKNOWLEDGMENTS

The authors are indebted to Professors Howard Aiken, Roman Jakobson, and Joshua Whatmough for their interest in this research and for many valuable suggestions. The work has been supported in part by the Harvard Foundation for Advanced Study and Research, the Ohio Oil Company, the United States Air Force, and the National Science Foundation.

REFERENCES

1. H.P.LUHN, "The Automatic Creation of Literature Abstracts," *IBM Journal of Research and Development, 2,* 159–165 (1958).
2. H.J.LANDAU, "An Automatic Glossary and Problems of Case Translation," *Paperis Presented at the Seminar in Mathematical Linguistics* (Vol. I), 1955 (On deposit at Wdener Library, Harvard University).
3. L.MATEJKA, "Selection and Classification of the Vocabulary for the Automatic Dictionary," *Design and Operation of Digital Calculating Machinery,* Progress Report AF-49, Sec. IV, Harvard Computation Laboratory, 1957.
4. D.M.BAUMANN, "A High-Scanning-Rate Storage Device for Computer Applications," *J. Association for Computing Machinery, 5,* 76–88 (1958).
5. G.W.KING, "A New Approach to Information Storage," *Control Engineering, 2,* No. 8, 48–53 (August 1955).
6. A.G.OETTINGER, *A Study for the Design of an Automatic Dictionary,* Doctoral Thesis, Harvard University, 1954.
7. K.MAGASSY, "An Automatic Method of Inflection for Russian," *Design and Operation of Digital Calculating Machinery,* Progress Report AF-49, Sec. III, Harvard Computation Laboratory, 1957.
8. V.GIULIANO, "Compilation of an Automatic Dictionary," *Design and Operation of Digital Calculating Machinery,* Progress Report AF-49, Sec. V, Harvard Computation Laboratory, 1957.
9. W.FOUST, "Inflected Form Generators," *Design and Operation of Digital Calculating Machinery,* Progress Report AF-49, Sec. VI, Harvard Computation Laboratory, 1957.

The Feasibility of Machine Searching of English Texts

VICTOR H.YNGVE

ABSTRACT Similarities between the literature search problem and mechanical translation raise the hope that an ultimate goal of directly searching texts written in English may be attainable. The role of a grammar in such a device is discussed. As a first step in reaching the ultimate goal, a very simple language is examined. This language, English Dialect A, has sentences consisting of single English terms. The hierarchical ordering of the terms is expressed in the grammar instead of in the terms. This provides certain advantages when used in a search program.

Literature searching and information retrieval problems in general and the Patent Office problem in particular are of interest because their solution would be of great practical value. They also raise questions of theoretical interest, the solution of which would advance greatly our understanding of human language and perhaps even of human knowledge. These problems concern the nature of the various artificial languages proposed for retrieval purposes and ultimately the nature of English, since it is from a background of English that these artificial languages are devised and since they are used for encoding information originally expressed in English.

The process of encoding information for retrieval purposes has similarities to other language translation processes which are being investigated intensively in the field of mechanical translation (MT). In this field, work is being done with the object of finding out how such natural languages as German, Russian, French, and English can be translated into one another automatically by machine.

The object of this series of reports is to see if some of the insights and techniques being discovered in the field of MT might be applicable to search and retrieval problems and to see if the insights developed by combining the two points of view might shed some light on the more basic linguistic problems.

VICTOR H.YNGVE Massachusetts Institute of Technology, Cambridge, Massachusetts, and supported in part by the National Bureau of Standards.

THE ULTIMATE SYSTEM

It is perhaps appropriate before embarking on a research program that we state an ultimate goal which is high enough to be a serious challenge and which can serve as a guiding principle during the course of the research. For the purposes of this series of reports we will take as our ultimate goal the development of a system in which the documents to be searched and the questions to be asked are expressed in English and all necessary operations are fully automatic. If we reach this goal we can say that the machine literature searching problem has been solved in principle. The word "English" will be used here to represent one of the natural languages. The ultimate solution could also be expressed in terms of languages other than English. The additional question of whether the proposed solution would be economically justified in any given situation can be answered more easily after we know how to search English text.

Recent developments in linguistics and in mechanical translation would lead one to expect that this goal may actually not be too remote to consider. It is the belief of some in the field of MT that it will eventually be possible to design routines for translating mechanically from one language to another without human intervention. Since accurate translating must leave the meaning unchanged but expressed in a different language and a literature searching operation must search for a particular set of meanings, there are many similarities between the two problems. However, since a searching system using English as the basic language has been explicitly rejected as too difficult and visionary if not actually impossible by some and merely overlooked or ignored by others, it might not be amiss to recount some of the obvious advantages of such a system before plunging into the not inconsiderable difficulties standing in the way.

The first advantage of searching English texts directly is that there would be no need for manually encoding the tremendous bulk of the patent literature, to say nothing of the other pertinent literature. It is of course assumed that a print reader can be developed which could serve as an input mechanism operating directly from the printed patent documents. The special problems associated with pictures and diagrams will not be discussed here. A second advantage associated with the elimination of the manual encoding step is the elimination of abstracting. The whole file would be available for search. Abstracting has the disadvantage that it must inevitably leave out some details as being unessential. These details then will not be encoded and can not be retrieved even if they are wanted at some future date. A third advantage is that the question posed by the examiner is already posed in English and he will not have to take the time to express it in a special machine language or wait for someone to do this for him. All communication between the machine and the patent examiner in this ultimate system would be in English.

The difficulties are not the language, but our understanding of it. The difficulties are not that English has no uniform or logical rule for the naming of things, not the ambiguousness of English words, not the wide diversity of phrasing and sentence structure which might be used in the same situation, not the arbitrariness of the conventions of language. English is in fact almost ideally suited to the search task. It is the language in which the patents are written and in which the questions are asked. It is the basis of the present patent classification system. It is used extensively in the present search procedure. Even all the decisions of the examiners as well as of the courts are based on how the patent document reads (in English), and on how the law reads (in English), and how it has been interpreted (in English). For all these purposes English serves us well. There are no difficulties with English, we use it effectively every day. The trouble is that we do not yet understand enough about the rules of our language to be able to instruct a machine to use it.

THE ROLE OF GRAMMAR

A language is a system of symbols and the rules for combining them which can be used for communication. The grammar of a language consists of a list of the symbols and a statement of the rules. Ultimately, a grammar will have to be contained in the machine if the documents and questions are available to it only in English. In MT we are finding out how to put grammars in a machine, but so far in information searching, little has been done along these lines. It may be of interest to see what the grammars of some of the machine languages in current use look like. Many of these languages consist essentially of descriptors. The grammar for a language like this is a list of the descriptors, a very simple grammar indeed, and the languages are very simple but apparently quite effective if properly used.

Most of the machine languages that I have seen proposed for search purposes seem to represent an attempt to find a happy compromise between two conflicting requirements. One of these requirements is that the language be simple enough to be used directly with searching devices which can carry out only simple operations like matching and elementary logical operations, for example, “indicate a match if you find descriptor A and descriptor B, but not descriptor C associated with the same document.” The other requirement is that the language be rich enough to express all the information through which it is desired to search, specifically information that is expressed in the document in English. Machine languages have been designed to attempt to reach an ideal compromise—to effect an impedance match between the English of the document and the binary decisions of the machine. One of the ingenious devices used is the method of showing explicitly in the code the inclusion relations between the terms. For example, the code for animal could be contained in the code for mammal and this in turn contained in the code for horse. This device is also used in the UDC. It has the advantage of reducing the grammar to a simple list and the search procedure to a simple match of a whole code word or part of one.

But the fact that a horse is a mammal and that a mammal is an animal is after all not a fact about the real world but a relation between symbols in our language. We just happen to have adopted the convention that certain animals are classed together and given a special term. The terminology is convenient but completely arbitrary. It is a part of the language, a fact of English grammar. If this fact is needed for search purposes, it should be stored in the machine as a rule of grammar, not furnished to the search mechanism each time explicitly in the code for horse. The problem is matching information originally expressed in English to the binary search criteria and binary search operations. It is asking too much to require that all possible answers to search questions be carried explicitly in the encoding language when they can more easily be carried implicitly in the language and brought into explicit form when needed by machine manipulation with the aid of a stored grammar.

A STEP-BY-STEP APPROACH

As important as it is to set a high ultimate goal, it is equally important to find a succession of short term goals, each of which can be quickly reached, each taking us one step nearer to the ultimate goal.

Many things will have to be discovered about English grammar before we will be able to search patents directly. For example, we will have to discover the various mechanisms the language uses to keep the reader informed as to whether the thing under discussion is the same thing that was mentioned before or something new. We will have to discover how a text can be unambiguous to the reader although nearly every individual word used is ambiguous in isolation. We will have to find out what is the connection between the subsumed-included relations familiar to the Patent Office and the linguistic categories familiar to those who have been working on the structural analysis of English. We will have to find formal connections between widely divergent ways of saying essentially the same thing. In addition there is much that we will have to learn about searching. If we had today a complete grammar of English which was capable of rendering explicit all the relations and distinctions implicit in the document, I doubt that we would know how to utilize it effectively in a machine search situation. We would be embarrassed by the very wealth of the information available. Much more must be learned about search situations.

As a first step in our approach to the ultimate goal, I suggest that we work with a very simple grammar which we will call English Dialect A. We will explore this dialect and its relations to the search problem carefully and learn from it what we can. Then we will devise an English Dialect B which will be more like English, and so on. These dialects will be chosen in such a way that we can reach an understanding of the search problem and of the linguistic situation in a relatively short time. At the completion of each step the dialect will be available for mechanization so that at any point a machine may be used to assist us in the research. Pilot or experimental systems can be set up at any point. Ultimately, we will be able to handle English as it is written; practical results applicable to particular search problems may appear from time to time along the way.

ENGLISH DIALECT A

English Dialect A is a language in which each sentence is one word. We will call the words terms, and list them in the grammar of the language. The terms are related only by means of a hierarchical system which is also expressed in the grammar.

The terms of English Dialect A are taken intact from English. They may be single words or phrases in English, but they are treated as single terms in Dialect A. The relation of Dialect A to English is that the meanings of the terms and their hierarchical relationships will be as near as possible to their meanings and relationships in English. Examples of terms in Dialect A are

COLLIE

CHESAPEAKE RETRIEVER

SHETLAND SHEEP DOG

HUNTING DOG

DOG FROM THE KENNELS OF JOHN SMITH

The rules of the grammar express the hierarchical system. They are formalized by writing pairs of terms in relations such as

ANIMAL= MAMMAL

MAMMAL=DOG

DOG=COLLIE

These rules are interpreted to mean that the term on the left represents a genus of which the term on the right is a species. It is clear that rules of this kind can completely specify a hierarchical system. It is also clear that the three rules above imply

ANIMAL	=DOG
ANIMAL	=COLLIE
MAMMAL	=COLLIE

so that it is unnecessary to write such rules in the grammar.

All the relations giving the species of one genus are collected together and called subrules of one rule. Since the left sides of these subrules are all the same, there is no ambiguity in omitting the left side for each subrule except the first one.

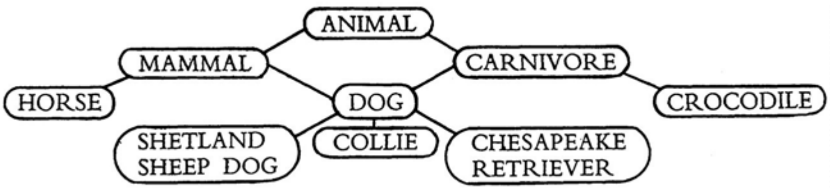
DOG	=COLLIE
	=CHESAPEAKE RETRIEVER
	=SHETLAND SHEEP DOG

It is also possible to have the same term subsumed under two or more genera. One may have for example

MAMMAL=DOG

CARNIVORE=DOG

The grammar thus does not exhibit a simple tree structure.



ANIMAL	=MAMMAL
	=CARNIVORE
MAMMAL	=DOG
	=HORSE
CARNIVORE	=DOG
	=CROCODILE
DOG	=COLLIE
	=CHESAPEAKE RETRIEVER
	=SHETLAND SHEEP DOG

A grammar can be interpreted as a computer program for deducing from a given term all the terms subsumed under it. The computer program would discover that subsumed under the term MAMMAL were to be found the terms HORSE, DOG, SHETLAND SHEEP DOG, COLLIE, and CHESA-PEAKE RETRIEVER.

Such a program would merely have to search among the terms on the left for MAMMAL, find that under MAMMAL are HORSE and DOG. The program would then search in turn for HORSE and DOG on the left hand side to find what is subsumed under them, and continue the process until the resulting terms could no longer be found on the left.

The grammar can also be interpreted the other way around, as a program for deducing from a given term all the terms under which it is subsumed. For this purpose, it is convenient to rewrite the grammar with the left and right hand sides of the rules reversed and the subrules reordered so that terms again appear only once on the left.

SHETLAND SHEEP DOG	=DOG
CHESAPEAKE RETRIEVER	=DOG
COLLIE	=DOG
DOG	=MAMMAL
	=CARNIVORE
HORSE	=MAMMAL
CROCODILE	=CARNIVORE
MAMMAL	=ANIMAL
CARNIVORE	=ANIMAL

We will call this latter a recognition grammar and the former a construction grammar.

A machine with a program of the above type could be used for literature searching in the following simple way: The documents would be represented by descriptors selected from among the terms of English Dialect A. An effort would be made to use the most specific terms possible for each document. The question would also be a term selected from those of the language, and would generally be a generic term. The program could operate on the descriptors of the documents using a recognition grammar, or it could operate on the question using a construction grammar, or both. In any case, additional terms would be generated for the document, for the question, or for both. The machine would then proceed to test for an exact match between a document term and a question term.

The next question to explore is what to do about ambiguous terms. For example,

MAMMAL	=DOG
DEVICE	=DOG

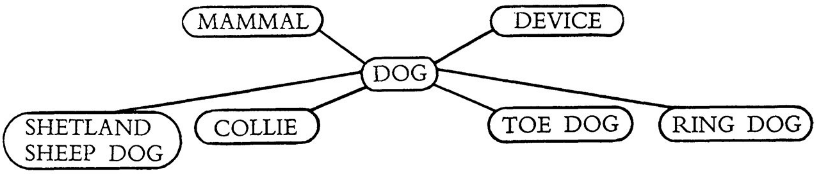
This situation is not to be confused with the previous problem

MAMMAL	=DOG
CARNIVORE	=DOG

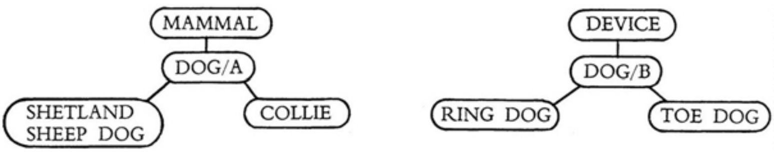
where DOG is really unambiguous. It has subsumed under it COLLIE, etc..

but not TOE DOG, RING DOG, or CHAIN RAFTING DOG, which will have to be subsumed under the DOG that is a DEVICE.

English Dialect A has no way of resolving the ambiguity of ambiguous terms, but English does, by the use of context, and dialects to be developed later will. But English Dialect A has a way of dealing with a problem arising from the use of ambiguous terms, that is, the problem of being able to deduce correctly that a COLLIE is a MAMMAL and that a TOE DOG is a DEVICE even though the language contains the ambiguous term DOG.



There are several ways of dealing with this problem. Perhaps the one best adapted to our purposes is a subscript notation. Two new terms are introduced into the grammar to resolve the ambiguity of DOG as far as the internal workings of the grammar are concerned.



ANIMAL	=MAMMAL
	=CARNIVORE
MAMMAL	=DOG/A
	=HORSE
DEVICE	=DOG/B
	=CLEVIS
CARNIVORE	=DOG/A
	=CROCODILE
DOG/A	=COLLIE
	=CHESAPEAKE RETRIEVER
	=SHETLAND SHEEP DOG
DOG/B	=TOE DOG
	=RING DOG
	=CHAIN RAFTING DOG

For the internal workings of the program, DOG/A and DOG/B are different terms, but for the purposes of comparison with descriptors or questions, the subscripts are ignored and DOG is ambiguous. It can only cause trouble now if it is actually used as a descriptor for a document or as a question. The trouble will take the form of selecting extra unwanted documents.

Our approach is not to try to invent unambiguous terms and require that the encoder and the questioner use them. Instead, our effort is to provide for the encoder and questioner a language that is as close to English as possible, so as to improve the match between man and machine. We arrange our grammar so that ambiguous terms, if used, will cause the minimum of trouble. The possibility of their incorporation gives us a language more like English and thus more natural to use. The ambiguity of a term is really not a property of the term, though we speak of it that way. It is the property of the grammar of the language to which the term belongs. A term may be ambiguous according to one grammar, but not according to another. Ultimately, we hope to have the grammar for a dialect of English that will effectively be able to handle the patents in their original language.

English Dialect A will be of interest linguistically if it is elaborated to include many more terms. Some of the formal devices that English uses to indicate class inclusion will become obvious through a comparison of the linguistic forms that English uses for generic and specific terms. English Dialect A may also be of some utility for immediate application to a certain class of information search problems.

ADDENDUM: THE SECOND STEP

REVIEW OF DIALECT A

English Dialect A was linguistically very elementary. It consisted of sentences that were composed of single terms. The terms were taken intact from English. They were noun phrases consisting of a noun head and one or more modifiers. The terms were arranged in hierarchical fashion by means of rules in a grammar which expressed the inclusion relations. The hierarchical structure was not a simple tree because a given term could be subsumed under more than one other term. Documents were to be encoded into terms of Dialect A and the questions were also to be posed as terms in Dialect A. But instead of searching by a simple match only, the machine would first generate all other relevant questions on the basis of the grammar and search for answers to them all. Alternatively, the machine could supply the documents with extra terms according to the grammar. Search was then to proceed on the basis of an exact match. It was conceived that any term used as a search question would retrieve any document described by that term or any term lower down in the hierarchy.

English Dialect A had important advantages over many of the other methods of encoding for search. The judgment of relevance was done by the machine on the basis of a stored grammar, not on the basis of the document codes alone.

The system could be used to find things from a completely new and different point of view merely by changing the grammar. No reencoding of the file would be necessary for this. The system could be brought up to date easily in the face of changes in the interests of the questioners. Older systems that attempted to incorporate the hierarchy directly in the codes had the disadvantage that the classification system could not be changed without reencoding, but in Dialect A, the classification system could be completely overhauled with no change at all in the encoding of the documents. In addition, the codes of Dialect A could be much more compact than codes carrying hierarchical information explicitly.

English Dialect A also had some serious shortcomings. It would have to contain a very large number of terms to be of much use, there being no facility to combine terms into more complicated expressions. No relations could be expressed between terms. There was also the problem that in Dialect A one always searched for a species, given a genus while, in fact, one sometimes wants to search for the genus given the species.

ENGLISH DIALECT B

English Dialect B bears some resemblance to English Dialect A. It has all of the advantages of Dialect A and some additional ones. Some of the shortcomings of Dialect A are eliminated in Dialect B. The main difference between the two dialects is that Dialect B has sentences consisting of two parts, a modifier and a term. The terms of Dialect B are the same as the terms of Dialect A, that is, they are nouns and expressions with noun heads and certain types of modifiers. By introducing a two part sentence, the number of possible sentences is not limited to the number of terms as it was in Dialect A, but approaches the product of the number of terms and modifiers. It is assumed in Dialect B that each modifier can be used with each term, an assumption that is not entirely warranted in practice.

A preliminary investigation was made of the requirements of a search and retrieval language, and of the kinds of simple dialects that seemed to hold promise of being useful. This investigation suggested that an understanding of certain special pre-noun modifiers would be a good second step in our understanding of how English serves as a retrieval language. Consequently, pre-noun modifiers were examined in some detail. Traditionally, pre-noun modifiers have been divided into two groups: the descriptive adjectives, such as *large, small, red,* and *old,* and the limiting adjectives such as *some, the,* and *these*. Of these two groups of adjectives, the descriptive adjectives have already been incorporated in the terms of Dialect A. It was felt that an investigation of their role in the retrieval situation could best be postponed until more was understood of the limiting adjectives.

It appears that the primary function of limiting adjectives is referential. They serve to refer to something (which is not directly named) in terms of some other thing or category (which *is* directly named). For example, the phrase *this dog* points out a certain definite object not named, but referred to in terms of a definite category designated by the term *dog*. These modifiers thus seem closely bound with the specification of sets and subsets of named items. Generally, the limiting adjective specifies the set or subset, and the descriptive adjectives and noun (a term in Dialect A) serve to name the set from which the subset has been taken. Of course, many if not all the descriptive adjectives can also serve to specify subsets. For example, *some tart apples* can be conceived of as a subset of *some apples,* but on the other hand, the set *some apples* is not guaranteed to contain such a subset. It seems best, therefore, to regard the descriptive adjectives as purely descriptive in line with the traditional view and treat them together with the noun as terms in a terminological hierarchy. In other words, the term *tart apple* is subsumed under the term *apple,* in the sense that all tart apples are also properly described as apples.

Limiting adjectives can readily be divided into two groups, those that render the noun phrases incorporating them self-contained so that their meanings are clear without reference to the immediate context, and those that require or imply reference to the context. *A, some, many* are in the first group. *This, those, the* are in the second group. Since our new dialect allows only one noun phrase to a sentence, it seems appropriate at this point to limit ourselves to the limiting adjectives that render the noun phrase, that is the sentence in our new dialect, self-contained.

Noun phrases in English can be classified into three mutually exclusive categories, those that are plural, those that are singular, and those that are uncountable. We have the contrast between *apples, an apple, apple*. Singular and plural are familiar enough. The uncountable category includes the so-called mass nouns (water), category names (sulphur), proper names (John), and so on. *Washington* in the sentence: “He lives in Washington.” is a proper name and therefore uncountable, although, of course, it can also be used in the singular and in the plural: “There are several Washingtons in the United States; the Washington that I mean is a state.” Almost all nouns can be used in all three categories. Some nouns undergo a definite meaning change when changing from uncountable to singular or plural: Carbon is an element, but a secretary makes a carbon of a letter. This phenomenon is ignored in English Dialect B. It should be investigated further in the future.

In the above examples, the three categories, plural, singular, and uncountable, are distinguished respectively by the plural *s,* the indefinite article *a* or *an,* and neither article nor plural *s*. We will thus take as three of the limiting adjectives in Dialect B, the following:

Ø — S
A — Ø
Ø — Ø

where we indicate zero by Ø to distinguish it from the letter O, and write A to stand for both A and AN. When these are combined with the term APPLE, we get

APPLES

AN APPLE

APPLE

We are now ready to specify in detail the structure of English Dialect B. The sentences consist of two parts, a term from English Dialect A and a modifier selected from the following list. To make a sentence, the terms are inserted in place of the X's in the list.

Ø X Ø
A X Ø
Ø X S
SOME X Ø
SOME X S
ALL X Ø
ALL X S
ANY X Ø
ANY X S
EVERY X Ø
MANY X S
ONE X Ø
THREE X S

as well as all the other numbers.

In order to be able to use Dialect B in a retrieval situation, we must know more of its grammar. Specifically, we must know how questions and answers are related in the language. For a question, we assume the following form:

Does this document show…?

and for encoded information, we assume the following form:

This document shows….

In the place of the three dots, we are at liberty to substitute any appropriate sentence from the dialect. In Dialect A, a document was found if it was described by a term lying lower down in the hierarchy than (subsumed under) the term used as the question. We now ask how this must be modified by the addition of the limiting adjectives. In order to obtain some information on this point, several questionnaires containing sample questions and document codes in Dialect B were circulated to Patent Office personnel. The results of these questionnaires have been carefully analyzed and have been partially incorporated in this dialect. Further investigations of this general nature may serve to modify the dialect in some details, but the overall structure will likely remain unchanged.

The first thing to be noted about the retrieval order is the behavior of expressions in the plural, singular, or uncountable. It turns out that a question in the uncountable should retrieve a document code in the singular and plural as well as in the uncountable. This is understandable because AN APPLE as well as APPLES contain the substance APPLE. This relation seems to be generally true for a large number of nouns. Also, the singular should retrieve the plural as well as the singular, since if one has APPLES, one also has AN APPLE in harmony with a broad interpretation of the meaning of the singular. We thus have in Fig. 1 the hierarchy of these sentences in Dialect B.

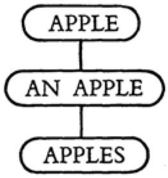


FIGURE 1. Retrieval diagram for plural, singular, and uncountable.

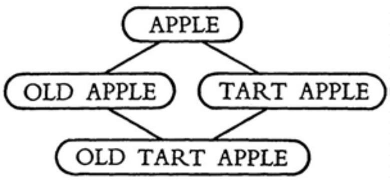


FIGURE 2. Retrieval diagram for some terms from Dialect A.

The next thing to investigate is what happens when there is a sentence with one term in the question and a sentence with another term, related to it hierarchically, in the descriptor. We have in Fig. 2 a hierarchy of terms of the type investigated in Dialect A. When each of these four terms is combined with the three modifiers, we have the result in Fig. 3. The correct diagram for

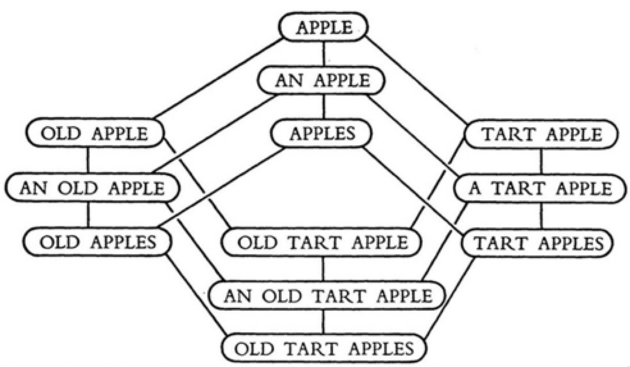


FIGURE 3. Retrieval diagram resulting from the combination of the diagrams of Fig. 1 and Fig. 2.

retrieval purposes is obtained by combining the first two diagrams in a way that one will recognize as the direct product.[1] In the diagram of Fig. 3, any one of the sentences may be used in a question. When so used, it should retrieve descriptors matching itself or any other sentence lower down in the diagram.

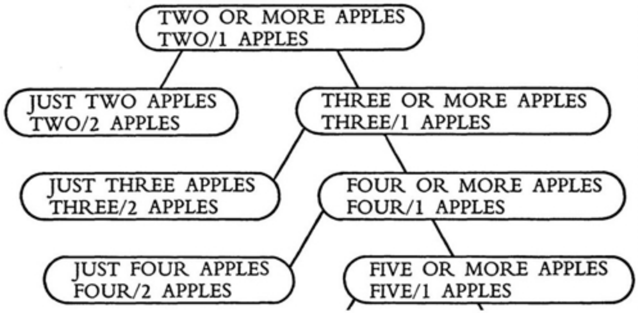


FIGURE 4. Retrieval diagram for sentences with numeral modifiers.

Investigation of the numerals as limiting adjectives reveals that in Patent Office practice they support at least two distinct meanings. THREE APPLES includes three or more apples, or just three apples. For our purposes, we can add subscripts on the numerals in the grammar as we did in Dialect A, and we can also add two new more precise modifiers and equate them to the subscripted numerals.

THREE/1≡THREE OR MORE

THREE/2≡JUST THREE

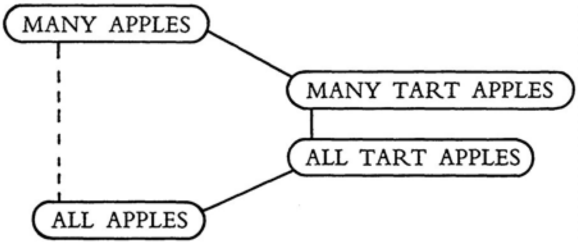
In line with the above considerations, it appears that the numeral plurals should be handled as shown in Fig. 4. When we take the direct product with the simple hierarchical diagram of terms given in Fig. 2, the result is rather obvious, but involved.

The results with the modifiers SOME and MANY show that for retrieval purposes we should set up the following equivalences:

Ø X Ø	≡ SOME X Ø		
A X Ø		≡ ONE/1 X Ø	≡ ONE OR MORE X S
Ø X S		≡ TWO/1 X S	≡ TWO OR MORE X S
	SOME X S	≡ THREE/1 X S	≡ THREE OR MORE X S
	MANY X S	≡ FOUR/1 X S	≡ FOUR OR MORE X S

[1] Calvin N.Mooers, "A Mathematical Theory of Language Symbols in Retrieval," page 1327.

The results with ALL X S and EVERY X Ø show that they should be considered equivalent in Dialect B for retrieval purposes. Furthermore, it is clear that they should be at the bottom end of the series of numeral adjectives discussed above, assuming that *all* is a great many. However, when one tries to take the direct product and draw a diagram of the retrieval hierarchy, one meets with a surprise. The question ALL TART APPLES retrieves the descriptor ALL APPLES. This is the reverse of all the other inclusions between TART APPLES and APPLES, and is shown in Fig. 5. It is clear that in the case of ALL, we cannot make use of the direct product as we did in the other cases.



Note: We assume that there are many apples.

FIGURE 5. Reversal of inclusion relations between sentences with ALL.

ALL is not the only modifier that exhibits this reversal of the hierarchical inclusion relations among the terms. ANY is another that behaves in this fashion, as in Fig. 6. There is another peculiarity with the word *any*. Of its several meanings, one can be used only in questions (and also in negative statements, but these are outside of Dialect B). In spoken English, the different uses of the word *any* are partly separated by stress:

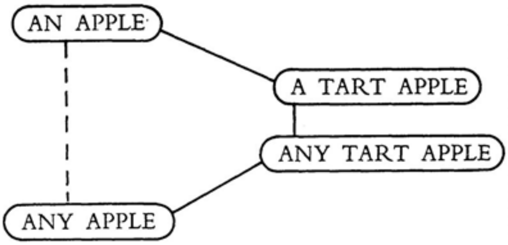


FIGURE 6. Reversal of inclusion relations between sentences with ANY.

Does this document show any_1 apples?
No, this document doesn't show any_1 apples.
Yes, this document shows an apple.
Does this document show $\textit{any}_2$ apples?

No, this document doesn't show *any*$_2$ apples, it shows some *tart* apples.

Yes, this document shows *any*$_2$ apples.

In its first use, *any* with the singular is used to question the uncountable category of mass nouns, whereas with the plural it is used to question the singular and plural (count nouns). In its second use, *any* is pronounced with stress and carries a meaning related to *any kind of, any species of,* or *any whatsoever,* again with the word *any* carrying stress. This is the use of ANY in Fig. 6.

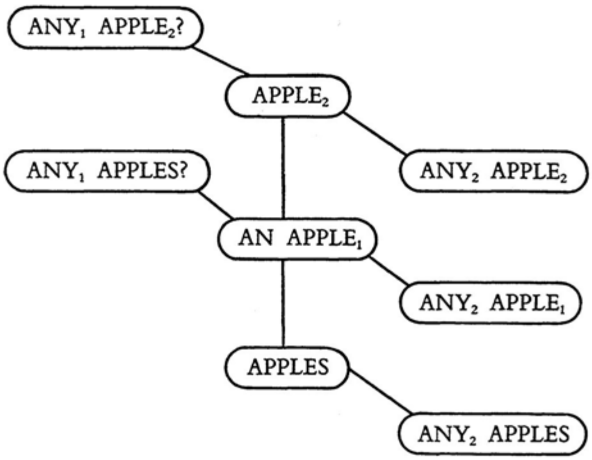


FIGURE 7. A retrieval diagram involving two meanings of ANY.

In Fig. 8, the phrases are from English rather than Dialect A, and various words are underlined to indicate where stress is to be placed when reading aloud. *Apple* has been subscripted to indicate the mass-count distinction in Figs. 7 and 8: subscript 1 for count singular, subscript 2 for mass.

Further investigation would be required to determine the degree of generality of these results. There may be classes of nouns that behave differently from *apple*. It is also not yet completely clear that the relation between APPLE and TART APPLE should be treated in the same way for search purposes as the different kind of relation between HALOGEN and CHLORINE.

The computer program for the grammar of Dialect B should be able to derive from the given question all the other possible implied questions that are needed for matching with the descriptors in the document, or it should derive possible questions from the document descriptors, or some combination of the two. Let us investigate only the former, deriving all descriptors that should be retrieved by the question.

We could, as we did in Dialect A, arrange the grammar in such a way that by moving down through the structure by a series of rules, one could come eventually to all of the points covered by the question. A better way, however, is not to work with the direct product, but to operate with two structures, one for each part of the sentence, the modifier and the term. This factorization of the product into modifier and term results in a great simplification in the program. One has a structure for the modifier in which one moves downward in order to find the points for which to search. At each point there is the possibility of choosing any of a series of terms from the hierarchy of terms. It is essential that the program go through the rules for the modifiers first before going to the hierarchy of terms because in some cases one has to move up through the hierarchy of terms, and in other cases one has to move down in the hierarchy. We will indicate this by giving the structure for the modifiers, and instead of using X to indicate where the various terms from the hierarchy should be placed in turn, we use A to represent a term and all those above it, i.e., more general, and V to represent a term and all those below it, i.e., more specific. A points up, V points down. We can then reserve the symbol X for those terms for which no other term in the hierarchy can be substituted. This has been done in Fig. 9. Modifiers involving ANY/1 can be used only in questions. All other modifiers can be used either for questions or for encoded document descriptors.

The grammar will be contained in the machine as a series of rules, much like the rules described for Dialect A. A possible method of search is then as follows. Locate the modifier from the question in the modifier structure and make a list of it and all modifiers below it in the structure. For each modifier in the list, make another list of it with all the terms (from the hierarchy of terms) that are either above or below the term in the question, as required, and then search the file for an occurrence of one of these derived descriptors.

There is, of course, the trouble that the list of modifiers is an infinite one, containing as it does all of the numbers. This and other problems affecting the speed of search can be handled by a slightly more sophisticated routine in a rather straightforward manner. Many of the techniques of search involving rules of progression, ordering, and screens, worked out for the HAYSTAQ system[2] can be used with advantage.

ASSESSING WHERE WE ARE

Let us examine Dialect B in the light of a clear and concise statement of objectives.[3] The first objective that is stressed is that a search is concerned much

2 B.E.Lanham, J.Leibowitz, H.R.Koller, and H.Pfeffer, Organization of Chemical Disclosures for Mechanized Retrieval, *Patent Office Research and Development Report No. 5,* U. S. Patent Office, June 14, 1957.

3 Don D.Andrews and Simon M.Newman, Activities and Objectives of the Office of Research and Development in the U. S. Patent Office, *J, of the Patent Office Soc., 40,* [2], 79–85 (1958).

more with interrelations between two or more elements than with the number of items or the detail with which they are found. This objective has not yet been met. It will be approached in later dialects. It seems necessary that we concern ourselves in the earlier dialects with how to describe and search for elements or items within the terminological structure of English before we tackle the problem of searching for them in combination or for their interrelations. It is worth mentioning, however, that we can already search in Dialect B for A CONNECTION, A MANUFACTURING PROCESS, A SUPPORT, THREE INTERMESHED GEARS, etc. We do have the disadvantage at present that we have to enter as terms, as in Dialect A, MANUFACTURING PROCESS, INTERMESHED GEARS, etc. We are not yet able to build these expressions up from their elements. This will come later, at the time when it seems to fit logically into a dialect and can be added easily.

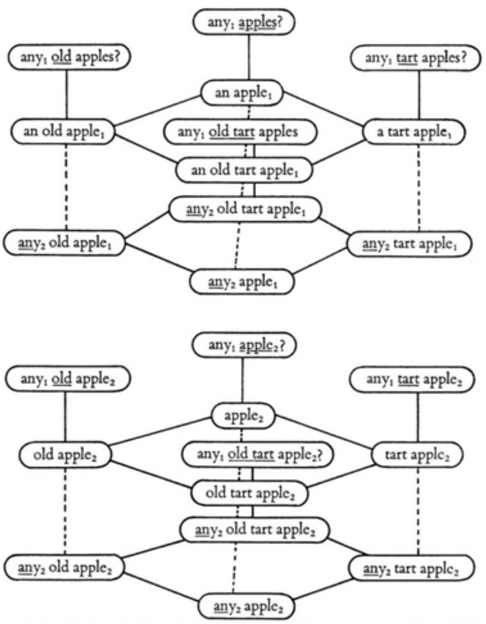
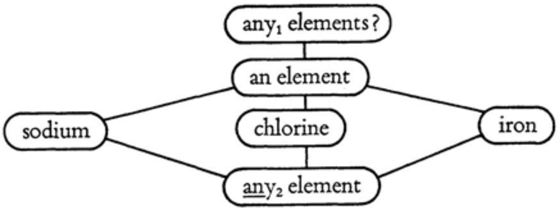
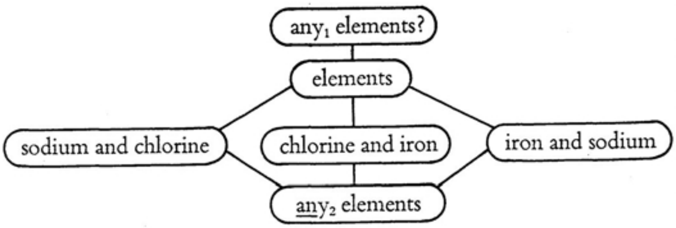
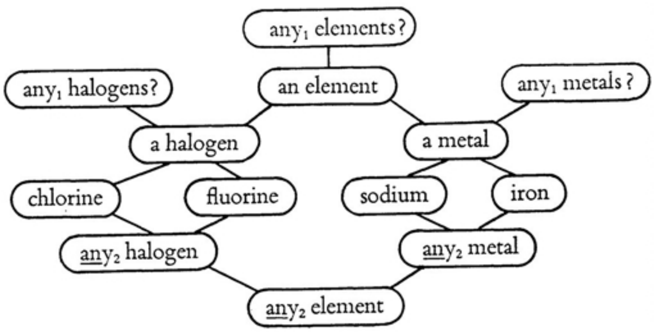


FIGURE 8. Retrieval diagrams involving the modifier ANY, and various hierarchical relations between the terms.







Even Dialect A had the capabilities of meeting the second objective, that every statement of a technical article should be retrievable in any frame of reference. The statements are searched for directly in their entirety. When new documents are added to the system it is only necessary to provide them with descriptors. In the ultimate system, the English of the document itself will be the descriptor for the document. The logic of relevance is determined on the basis of the grammar in the machine, and not on the encoded form of the descriptors. For this reason, it is very easy to incorporate the developing experience of the users of the system. In order to expand or change the logic of inclusion that the system operates with, it is only necessary to modify the grammar in the machine. One does not have to alter the codes associated with the documents.

The third desired feature of patent application searching was the ability to retrieve a species when a genus is requested, and also, where applicable, retrieve a genus when a species is requested. This requirement was not met in Dialect A, but it has now been met in Dialect B with its inclusion of ANY, EACH, and ALL.

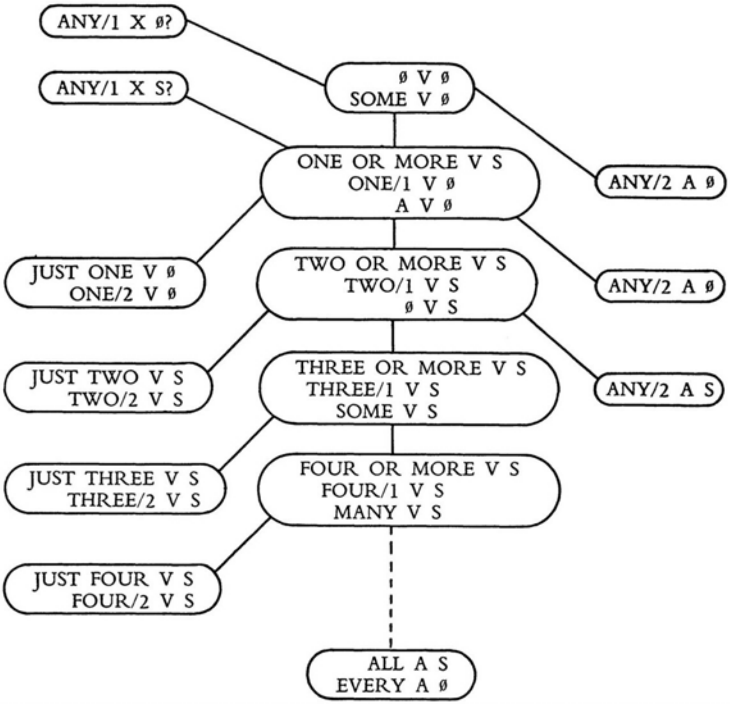


FIGURE 9. The final retrieval diagram for sentences in Dialect B.

The fourth and fifth features, ability to recognize alternative items and the ability to handle implicit or explicit absence of features have not yet been included.

Probably the most important problem remaining is the first one, representing interrelations. Our procedure in deciding what to add to a dialect to get another dialect is to try to add what seems to be most needed. If it cannot be added because the dialect must have something else first, then we try to add the prerequisite. The final decision can be considered as a compromise between what is most needed and what can be added most easily at that stage.

REMAINING PROBLEMS

Many questions have been raised by this investigation that should be answered. Some have already been mentioned; some others are listed here. Every question that is answered satisfactorily will take us just that much closer to our ultimate goal of being able to search English texts directly.

1. What other modifiers in English share with the modifiers of Dialect B the property of being clear without reference to the immediate context?
2. Which of these behave in an identical fashion for search purposes to the ones treated here?
3. What is the behavior of the others in the retrieval situation, and how can they be incorporated in a dialect?
4. Are there classes of terms that would require a different hierarchy of modifiers?
5. Can the meaning changes of some terms when used with different modifiers be systematized?
6. What would have to be done to introduce the descriptive adjectives into a dialect separately from the nouns so as to reduce greatly the number of items that would have to be stored in the grammar?
7. Are the inclusion relations between the same noun with different descriptive adjectives independent of the choice of noun?
8. Are the inclusion relations of different nouns with the same descriptive adjectives independent of the choice of adjective?

Semantic Matrices

G.PATRICK MEREDITH

ABSTRACT. The semantic matrix is a grap hical device for plotting in a standard conventional form whatever precise elements of meaning have been ascertained from the semantic analysis of a concept.

The intended use of the device is to provide standard descriptions of the structure of conceptual stimuli in order to facilitate the comparison and replication of researches on concept-formation, communication, and comprehension.

The device may thus be compared with the standard Cartesian graph-convention for translating algebraic functions into geometrical forms The. notion may be regarded as a development of the “matrix-function” introduced by Whitehead and Russell in the *Principia Mathematica*. This was conceived as an array of propositions. The semantic matrix is a generalization of this.

Any array of symbols having determinate meanings whose interrelations are expressed by spatial positions may be said to constitute a semantic matrix. Pascal's triangle of binomial coefficients is an example. But we are not restricted to numerical elements. The elements, however, must be clearly defined, both as regards semantic content and as regards syntactic function. This approach to the meaning of conceptual words reveals that the comprehension of such words often entails the grasp of a considerable wealth of essential implications. The semantic analysis may result in a hierarchy of conceptual elements, analogous to the Fourier series resulting from the analysis of a complex wave-form. This richness of meaning is described as “benign ambiguity.”

The formal theory developed in the present paper defines the *referential elements,* the *syntactic functions,* the *matrical structure,* and the *principal variables* in the design of semantic matrices. An *inductive theory* can be developed from a survey of existing graphic devices but this will demand prolonged research. A *constructive theory* is here presented in which certain arbitrary conventions are set up and their use is indicated.

In the field of scientific communication we have to deal not only with communication between experts or communication between the expert and the public (i.e., popularization) but also between experts in different fields, unfamiliar with one another's concepts. In an age of interdisciplinary projects this last type of communication is of especial importance. The Cartesian graph provides one invaluable medium of interdisciplinary communication, but it is restricted to certain types of meaning only. The

G.PATRICK MEREDITH Department of Psychology, The University, Leeds, England.

constructive theory here offered deals first with some of the contextual problems arising in the analysis of concepts and goes on to establish logical conventions for the construction of matrices.

When logic is transposed from the context of philosophical discrimination to one of practical communication, new insights appear. In particular the traditional subject-predicate dichotomy is found to have a new psychological and linguistic significance. Again the class-calculus, Euler's circles, and the theory of sets, when stripped of non-essentials, yield a simple graphical device very similar to a Cartesian graph in which an unlimited range of syntactic forms can be expressed, once the conventions are grasped. Each of these forms reveals a certain “inferential potential” which is of direct importance in comprehension and communication.

The conventions used here are based on a system of logical forms constituting a logical calculus designed for the analytical problems arising in research on the comprehensibility of scientific and technical reports and published in a separate paper (1).

THE CONTROL OF AMBIGUITY

CONCEPTUAL AND INSTRUMENTAL FORMS

The concept of the *semantic matrix* is a development of the logical matrix as originally presented by Whitehead and Russell in *Principia Mathematica*. Quoting from the Introduction to the First Edition, 1910:

When something is asserted or denied about all possible values or about some [undetermined] possible values of a variable, that variable is called *apparent,* after Peano. The presence of the words *all* or *some* in a proposition indicates the presence of an apparent variable; but often an apparent variable is really present where language does not at once indicate its presence. Thus for example “A is mortal” means “there is a time at which A will die.” Thus a variable time occurs as apparent variable.

Whatever may be the instances of propositions not containing apparent variables, it is obvious that prepositional functions whose values do not contain apparent variables are the source of propositions containing apparent variables, in the sense in which the function φx is the source of the proposition $(x) \cdot \varphi x$. For the values for φx do not contain the apparent variable x, which appears in $(x)\ \varphi x$; if they contain an apparent variable y, this can be similarly eliminated, and so on. This process must come to an end, since no proposition which we can apprehend can contain more than a finite number of apparent variables, on the ground that whatever we can apprehend must be of finite complexity. Thus we must arrive at last at a function of as many variables as there have been stages in reaching it from our original proposition, and this function will be such that its values contain no apparent variables. We may call this function the *matrix* of our original proposition and of any other propositions and functions to be obtained by turning some of the arguments to the function into apparent variables. Thus, for example, if we have a matrix-function whose values are $\varphi\ (x,y)$, we shall derive from it

$(y) \cdot \varphi\ (x>y)$, which is a function of x,

$(x) \cdot \varphi\ (x,y)$, which is a function of y,

$(x,y) \cdot \varphi(x,y)$, meaning "$\varphi(x,y)$ is true with all possible values of x and y." This last is a proposition containing no *real* variable, i.e., no variable except apparent variables.

It is thus plain that all possible propositions and functions are obtainable from matrices by the process of turning the arguments to the matrices into apparent variables.

There are two points to note about this account. The first is that the matrix is conceived as a *source of propositions*. It is a condensed symbolism from which, by following certain rules, a series of consequential propositions may be obtained deductively. We may compare it with an algebraic matrix, say

$$A=a_{ij}$$

from which, by giving i and j respectively all the values from 1 to m and 1 to n, we obtain the extended form

$$\begin{Vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & & a_{mn} \end{Vmatrix}$$

The second point is that whereas in algebra the presentation in the extended form is often deemed essential, the prepositional matrix is nowhere presented in extended form in the *Principia*. I shall distinguish these two ways of using matrices as the "conceptual use" (non-extended) and the "instrumental use" (extended). For it is by spreading the elements of the matrix out in space that we are enabled to develop the consequences of their mutual relations. The spatial extension functions as an instrument of thought.

Later authorities on logic have indeed made some small use of extensional forms, the most obvious example being the truth tables introduced by Wittgenstein. This is a clear illustration of the fruitfulness of extended matrices as instruments of thought, for these tables enable the truth or falsity to be determined for logical propositions involving combinations of logical constants to any degree of complexity.

An extension of this technique is given by Bochenski (6) in a conveniently succinct notation for the logical constants. For example *implication* is symbolised thus:

⊃	1	0
1	1	0
0	1	1

Here the truth-values in the lower right-hand quadrant are each a sort of product of the values in the row and column headings. It is as if we said:

1×1=1	1×0=0
0×1=1	0×0=1

(This is the peculiar arithmetic of “material implication.” For the proposition “*p* implies *q*” is false only for the case when *p* is true and *q* is false, i.e., 1×0= 0.)

OPERATIONAL USE OF SPATIAL RELATIONS

The purpose of these examples is to stress the instrumental value of the spatial extension of symbols. From this point of view we may regard Descartes and Cayley as being complementary influences in the history of mathematics. For whereas Descartes related *quantity* and *space* through “algebraic geometry,” we might say that Cayley completed the relation by inventing what is essentially a “geometric algebra.” Many mathematical expressions, e.g., Pascal's triangle, gain in significance when we perceive that the spatial arrangement-endows the geometric relations of the elements with a specific operational connotation. For consider the first few rows of the triangle:

```
      1   1
    1   2   1
  1   3   3   1
1   4   6   4   1
```

The rule is that the sum of any two adjoining elements yields the element on the next line under the point midway between them. Clearly we are free to attach any rule we please to any spatial relation. We can thus generate an indefinite variety of extended symbolic forms. The algorithms for the four fundamental operations of arithmetic can be viewed in this light, as can the rules for vulgar fractions. There is, in fact, a kind of tacit *spatial grammar* at work throughout the whole of mathematical symbolism. We do not need to go into two dimensions to appreciate this. The simple algebraic symbolism for a product, *xy,* is a tacit use of linear juxtaposition to symbolize an operation. But of course the power of the matrix notation lies in the greater range of spatial relations which are opened up by a two-dimensional arrangement.

GENERALIZATION OF STATUS

But whereas in algebra the elements of a matrix are, in general, deemed to be *quantities* of some type or other, or at any rate quantitative operators (such as differentials), the propositional matrix of the *Principia* points to a much more general concept. The elements need not be quantitative. They can have any logical status whatever. And the spatial relations between elements can be associated with any rules of operation to suit our convenience. It should be evident that we have here a symbolic device of exceptional generality and power awaiting precise development. For insofar as any concept has a logical structure which can be made articulate, i.e., expressed as a group of logically related elements, it can be represented as a matrix. And by the fact of possessing a defined logical status, the matrix can enter into the operations of a logical calculus. It is to this general concept that I have assigned the name "Semantic Matrix."

COMMUNICATION OF SCIENTIFIC CONCEPTS

On the mathematical side a development analogous to that of the tensor calculus may be envisaged for these functions. But in view of the endless variability in the logical status of their elements (and hence in each type of matrix as a whole) some pragmatic criteria may be desirable in order to ensure that the development does not become a trivial proliferation of pretty patterns. "Pragmatic" can here be taken in two senses, viz., (1) criteria of relevance to the systematic development of symbolic logic *or* (2) criteria of relevance to the manipulation of concrete concepts such as those occurring in science. Both developments are desirable, but in this paper I shall pursue the second since the concept of the semantic matrix did, in fact, arise in a context of practical research on the communication of scientific concepts.[1] A word or two on this problem will indicate the need for such a technique as that provided by the semantic matrix.

It is popularly supposed that the main reason for the difficulties encountered by the non-specialist seeking to understand a scientific document is the scientist's use of "jargon." Now it is true that scientists whose daily preoccupation is with physical problems do not always display the same facility with words as some of their colleagues for whom words are the stock-in-trade. But this is surely only a minor and remediable source of difficulty. The major source lies in the inherent complexity of the concepts to be communicated. Often an explanation in simple words, whilst theoretically possible, can be achieved only at the price of such prolixity as to defeat the ends of the explanation. (The analysis of a single word often occupies, for example, a paper of five to ten thousand words in the *Proceedings of the Aristotelian Society*.)

ANALOGY WITH RADIO COMMUNICATION

Any attempt to achieve an adequate solution of the problem of communicating scientific concepts through mere verbal simplification must in the long run prove abortive. I found a more radical approach imperative. If we con

[1] In a research in the University of Leeds, Department of Psychology, on the *Comprehensibility of Technical Reports* sponsored by the D.S.I.R. under the Conditional Aid Scheme 1953. For the development of the mathematical aspect of this research see *Epistemic Communication,* Part I, The Modular Calculus, *Proceedings of the Leeds Philosophical and Literary Society, 1958*.

sider the partially analogous problem of radio communication we get some hint of a possible solution. The solution was in two stages: (1) the understanding of the nature, structure and control of electromagnetic radiation, and (2) the design of transmitters and receivers to permit resonance in the latter to the controlled emissions of the former.

In our problem we have, corresponding with the flow of radiation, a flow of scientific concepts emanating from the scientific journals. Our first task is therefore to understand the nature, structure, and control of this flow. For this we need something analogous to Clerk Maxwell's equations. The analogy may here be pursued even more closely (bearing in mind the dangers in all analogy).

The inseparability of electric and magnetic occurrences and the curiously reciprocal relations between them formed the substance of Faraday's fundamental discoveries and were subsequently formalized by Clerk Maxwell. There is likewise a duality in every scientific concept. On the one side it refers to concrete experiences and on the other to logical categories. Indeed this is true of ordinary words in a sentence. Every word has a “meaning” to convey and every word likewise exerts some grammatical function in the sentence. And these two, though different, are inseparable, except through a process of abstraction. Through abstraction logic can deal with the functional aspect of words apart from their meanings, and semantics can deal with their concrete references. But if either discipline strays too far from the other it tends towards triviality. At the same time we have as yet no adequate “logic of intensions” by which the two can be functionally joined, and logicians display an understandable reluctance to sacrifice their freedom in the realm of abstraction in the interests of concrete interpretations. Approaching the duality from the other side, viz., from the problem of communicating concepts I see in Russell's propositional matrix the first hint of a functional bridge between logic and semantics. The semantic matrix represents an underpinning of this bridge to permit the passage of traffic.

THE CONTROL OF AMBIGUITY

Ambiguity plays a little-understood role in scientific communications. It may take two forms, “benign” or “malignant.” In the benign form it adds to the overt meaning of a term a succession of underlying layers of meaning which amplify and enrich the overt meaning. The depth of meaning to which any individual reader can penetrate is a function of his own conceptual equipment. Malignant ambiguity is the more usually understood sense of the term. Here a confused choice between alternative and independent or even incompatible meanings confronts the reader. When a concept can be expressed as a semantic matrix the malignant ambiguity is minimised and the benign ambiguity is given overt expression.

Again using analogy we may compare Fourier's approach to the flow of heat[2] in which earlier attempts by Lagrange and Dirichlet to express functions as sums of sinusoidal components were superseded by a quite general theorem thereby establishing the so-called “Fourier series” now in universal use. In our present terminology the coefficients of a Fourier series may be regarded as the overt expression of the benign ambiguity latent in the original function. What we seek in the semantic matrix is the equivalent in logical categories of the expression of successive refinements of algebraic “meaning” in a complex function.

FORMAL THEORY OF SEMANTIC MATRICES

THE ELEMENTS

Definitions

1. A *semantic matrix* is a display of referential elements.
2. A *referential element* is a material sign conveying a meaning to a human respondent.
3. A *display* is an arrangement of elements in a spatial framework according to some ordering principle.
4. Semantic matrices *develop functional properties* in so far as their elements are assigned characteristic syntactic functions.
5. The *syntactic function* of a sign is a rule of usage whereby the sign enters into characteristic relations with other signs to establish a matrical structure of signs.
6. The *matrical structure* of a semantic matrix enables it to enter into inter-matrical relations with other semantic matrices.

Examples

7. An ordinary grammatical sentence is a semantic matrix. The words are the referential elements. The parts of speech are the syntactic functions. The linear arrangement is the display. The ordering principle is determined by the rules of grammar.
8. Maxwell's colour triangle is also a semantic matrix. Here the referential elements are points either on the periphery or within the interior. These points may or may not bear labels. They refer to colours formed by mixing three primary hues in differing proportions. The syntactic functions of all the points may be described as “adjectival” since each refers to a colour-quality. The triangular arrangement is the display. The ordering principle is the relation between *position* and *proportion*.

[2] *Théorie analytique de la chaleur,* 1822.

Comparison of examples

9. The first display is a linear array; the second is two-dimensional. The first set of elements exhibits a mixture of syntactic functions—it may be called “hetero-typic.” The second contains elements all identical in function: it is “homo-typic.” In the first the ordering principle is grammatical. In the second it is mathematical.

Further examples

10. The Periodic Table of Elements. A family tree. A paradigm for the declension of a Latin noun. A determinant.

Comment

11. It will be seen that semantic matrices vary over a wide range of instances. But in every instance we have a display of meanings compactly expressed and arranged in such a way as to exhibit systematic relations. The properties of such displays, both formal and functional, merit a thorough analysis. Such an analysis would provide a basis for a general grammar of sign-systems. It would also facilitate the consistent treatment of many problems of communication. The above examples concern matrical structures only, not inter-matrical relations. The latter cannot be discussed profitably until we have formalized *intra*-matrical syntax.

PRINCIPAL VARIABLES

Definitions

1. Every semantic matrix has four characteristics:

(i) *Constitution,* i.e., the visual appearance of the constituent elements,
(ii) *Structure,* i.e., the spatial arrangement of the elements,
(iii) *Functional Character,* i.e., the totality of syntactic functions exhibited.
(iv) *Reference,* i.e., the field of meanings involved.

Prescription

2. Since the matrix is a spatial display the primary character of the elements is visual. Whenever possible every element is to be assigned a *phonetic rendering* to facilitate communication. In the translation suitable cues to spatial structure are to be provided.

Example

3. x^5 is a simple visual matrix. The same, read aloud as “ex to the fifth” is

the phonic counterpart. The words "to the" give the cue to the position of "5."

Prescription

4. The visual elements may take any form whatever—pictorial, hieroglyphic, lexical, symbolic, geometrical, including mere points. A reason should always be given for the choice of form.

Prescription

5. In the spatial arrangement of the elements the geometric relations should be correlated with the logical relations between the elements. This applies not only to juxtaposed elements but ideally to any two elements.

Example

6. A family tree well illustrates the exact correlation of geometric with logical relations. The systematic family connexion between any two persons in the tree is immediately deducible from the sequence of vertical and horizontal steps. But it must be observed that such deductions are not solely a matter of geometry. The sex and marital status of each person are involved in the deduction.

Prescription

7. Elements are not, in general, mere points having nothing but geometrical relations with one another. They are symbols having not only semantic reference but a certain syntactic function.

Discussion

8. The theory can develop in two directions: (i) towards a systematic analysis of existing usage, and (ii) towards a system of rules for constructing semantic matrices for specified purposes. Since we are working towards a novel conception, viz., a *geometrical syntax,* whose working is not easy to anticipate, it is advisable to work empirically at first. Innumerable examples of tabular and other non-linear arrangements of words and symbols already exist. Grammarians have attended exclusively to the linear arrangement of words in sentences. This conventional grammar must now be regarded as a particular case of a very much more extensive system, just as Aristotelian logic has turned out to be a special case of the much wider system of symbolic logic. The richness of the field to be explored can be illustrated by the Periodic Table of Chemical Elements. The power of this visual display in generating systematic inferences concerning relations between its constituents indicates the latent potentiality of our nascent geometrical syntax.

APPLICATION OF THE PRINCIPLES

We come now to what is perhaps the central principle in the theory of semantic matrices. It must be borne in mind that their primary function is to facilitate understanding. To understand a concept is to apprehend correctly all the relations which determine its structure. This means not only grasping the fact that certain relations hold between certain terms but also seeing that the nature of the terms permits those relations to hold and that the global character of the concept determines their occurrence. For example, to understand the ecology of a hedgerow we have to understand that a whole system of predatory, parasitic, nutritive, and other relations hold between the various species of organisms in the hedgerow; that the nature of these organisms permit these relations and that the size, situation, and climatic exposure of the hedgerow determine the persistence of these relations, so that an ecological balance is maintained. This is a familiar enough concept for the trained biologist, but the function of semantic matrices is to analyse the problem of the learner faced with unfamiliar concepts.

A caterpillar, a hawthorn, a bird, a micro-organism, such may be the familiar *subject-matter*. But how familiar are they? Each has thousands of genes conspiring to produce innumerable characters. Through these characters the organism enters into relations with other organisms. To be "familiar" with a caterpillar may be no more than to recognize it at sight. To understand its role in an ecological system demands more than this. We cannot simply express the relation between two organisms, as the logicians do, by "x R y." We have to find expression for what it is in x and in y which enables R to hold between them. We are not concerned with merely asserting a relation but with communicating an understanding of how the relation comes to hold. This demands that the meanings of x and y be expanded sufficiently to indicate the basis of the relation. *The function of the semantic matrix is to supply this expansion of meaning.*

As already indicated we can regard any visual expression of meanings as a semantic matrix but obviously some are much more explicit than others in their representation of meaning and structure. The position, then, is that countless semantic matrices already exist, though unnamed as such, but their geometric syntax (i.e., the logical relations expressed in the relative spatial dispositions of the elements) is commonly unstated and unexplored. The reason for this neglect is not far to seek. This is not a problem in pure mathematics, or in grammar, or in logic. It is one which necessarily draws upon all three and more besides. A new and even more resourceful George Boole is needed to delineate the cartography of this new domain of logic. It had better be a young man for much fresh thinking is needed and the territory will need a lifetime to explore.

In a short introduction I can only indicate what seem to me the two main directions which developments must take. I shall call these the inductive theory and the constructive theory, respectively. The first sets out to explore the existing forms of semantic matrices and to establish what may be called their “natural history.” From this may be derived an empirical taxonomy. Thence the inquiry will branch into two directions. One will seek to formulate the principles of geometric syntax implicitly and intuitively employed by the authors of the existing matrices, and to separate those which do display logical order from those which do not. The other, taking the matrices as communicative devices which have to be perceived, interpreted, and comprehended, will investigate, by the methods of experimental psychology, the cognitive processes by which these matrices convey their meanings. We know something of the perception of geometric forms and something of the psychology of reading, but what hidden principles are at work in the interpretation of symbolic forms spread out in geometric arrangements? Such will be the quest in the inductive theory.

In the constructive theory we shall start not from existing forms of representation but from logical forms. Portions of knowledge to be communicated are translated into these logical forms and then arranged geometrically according to certain arbitrary but systematic codes. The knowledge is also expressed in conventional textual fashion. Experiments in comprehension are then carried out. The method, briefly, is as follows: a sufficiently large group of suitable readers is split in half. The topic to be communicated is expressed both conventionally and in a document based on analysis by semantic matrices. Both versions are split in half. One version is presented to one half-group of readers, the other version to the other half-group. For the second half of the topic the versions are changed over so that each half-group is “exposed” to both types of version. The tests of comprehension and the methods of scoring are both of unusual design, specially adapted to yield the information desired.

Briefly the information we seek in these experiments is this: which kinds of knowledge are more readily communicated conventionally and which kinds by documents derived from semantic matrices? There are, of course, many kinds of textual expression and many kinds of geometric arrangement and hence an extensive programme of experimentation is required. In the end we have to compare the *best* we can do in textual expression with the best in semantic matrices before deciding that one or other is superior, and even then this superiority refers to the communication of only certain *kinds* of knowledge.

The motivation of this development lies in the need to improve our methods of communicating scientific knowledge, whether in education, in technology, in scientific journalism or in research. The impact of automation and its associated developments on our patterns of occupation and social organization is so forceful that unless quite radical changes in our methods of communication can be introduced only a technocratic dictatorship will be able to hold society together. This is not a matter of improving our technical media of communication, indeed it is these which are setting the pace. It is in the processes of interpretation and comprehension that our problem lies. It is useless to step up the bombardment of the human organism by pellets of information unless the pellets are so organized as to be capable of faster assimilation.

We have hitherto failed to correlate two fundamentally distinct modes of apprehension, viz., the linear reception of linguistic information in time and the two-dimensional reception of structural information on a surface. The fact that scanning-processes can convert the latter into the former must not blind us to the essential differences between the two modes of perception. It may well be that linear information, after reception in time, becomes spread out in cerebral space. It is true that given a well-learnt diagram we can wander about in it mentally building up our understanding by detecting its internal relationships. We can even reconstruct such a diagram from verbal descriptions. *But are these processes efficient?* Every shift of the focus of attention demands a fresh effort of recall. It is not in the nature of the brain substance to keep all the parts of a cognitive pattern simultaneously at the same level of vividness. But ink on paper preserves its sharpness indefinitely.

The above indicates the limitations of purely oral exposition. In a somewhat lesser degree it applies to verbal expositions presented textually. For although a printed page is two-dimensional, the printed language is normally presented in *lines* of type. It is true that very expert readers are able to take in the sense of a paragraph or even a whole page apparently at a glance, but this accomplishment is not widespread among those most in need of aids to comprehension. It is probably effective only with relatively familiar subject-matter and is not likely, for example, to operate for a physicist seeking a rapid grasp of the principles of genetics. And today many of the acutest problems of communication arise in the fields of inter-disciplinary research.

The development of semantic matrices is necessarily based on mathematical analogies. A caution is needed here. Analogies are notoriously liable to misuse. Yet the whole development of science and mathematics has rested on the triumphant exploitation of analogy. We shall be dealing with qualitative variables akin to the terms of symbolic logic (but with an intensional reference) more often than with the quantitative variables of elementary algebra. But we shall speak of *functions,* and here it should be pointed out that for many purposes, even in algebra, functions are dealt with without any guarantee that their equations are known. We must not identify the function with its equation. A function expresses a certain distribution of values. The equation specifies the exact relation of dependence between the dependent and independent variables. The function exists as an empirical fact whether we can write down its equation or not. Thus in speaking of functions we are referring to patterns of values of variables without having to commit ourselves to the possibility of algebraic treatment. Indeed for many of the problems of communication the latter would be irrelevant. But it also follows that in any mathematical analogy we may use we must guard each step against any illicit importation of purely algebraic implications. What we seek to follow here are the *strategies* by which the great algebraists of the 19th century, Galois, Hermite, Kronecker, Cayley, Sylvester, Hilbert, and the rest, transformed their science, by attention to topological and other non-quantitative relations. And since we are dealing with problems in the mapping of knowledge, we shall also draw upon the geometric insight of Riemann and the automorphic functions of Poincaré.

Semantic matrices form part of a wider discipline, viz., the modular calculus which deals with the mathematical problems of epistemic communication (1). An outline of the concepts which have guided these developments is needed here in order to indicate the reason for the special character of semantic matrices.

The standard situation from which we start is that of a reader studying a scientific document. What are the factors determining his understanding? The document, in general, is about certain objects, organisms, instruments, etc., and may give an account of certain phenomena associated with these objects, certain laws, principles and theories, and the vocabulary and symbols required for the exposition. Some of the material of the document must be familiar to the reader, otherwise he has no foothold, e.g., certain common properties of objects and some of the vocabulary. On this basis of familiarity (the "subject-matter") the writer has to construct unfamiliar concepts (the "predicate-matter"). Obviously there is no problem of comprehension unless some of the material is unfamiliar. For example the subject-matter might be a set of radio-components and the predicate-matter might be a radio circuit.

We assume that the reading is purposive. The reader wants to master the predicate-matter, either in order to construct the radio-set or to improve his knowledge of radio-theory, or both. To determine whether the document is successful in fulfilling its communicative function we must therefore submit the reader to a practical test or a written test or both. Obviously his understanding may vary considerably in both depth and extent. Our design of the test will be an expression of the kind of understanding we want for him.

This whole situation is schematized in "The Square of Communication," Fig. 1, presented in a symposium on semantics at the British Association for the Advancement of Science, Bristol, 1955 (2). For our present purpose the square may be elaborated and modified a little (Fig. 2). The original square was a development from the famous triangle of Ogden and Richards (3).

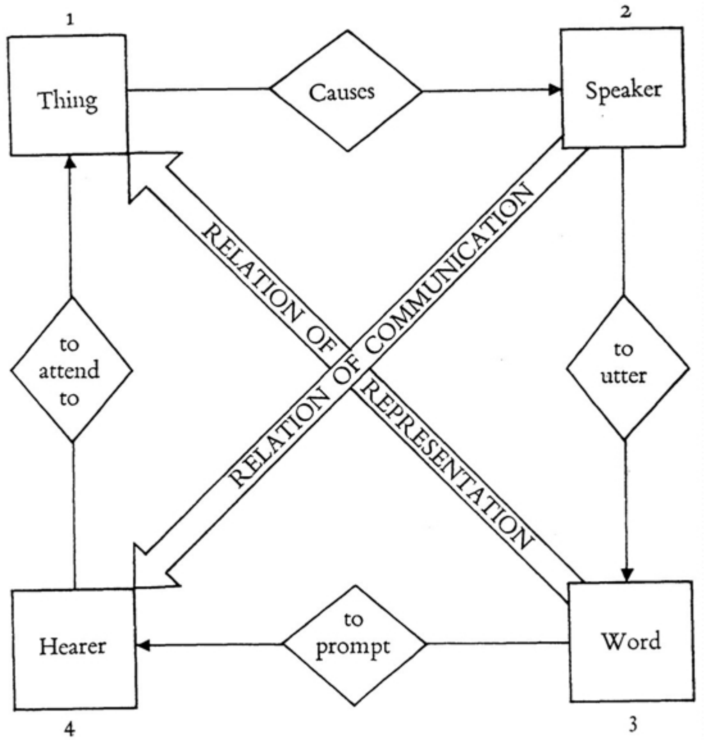


FIGURE 1. The square of communication.

Now it has been assumed that the writer of the document is also the investigator or inventor of the original predicate-matter. But often this is not the case and, in general, it is better to introduce another agent into the situation. For the specialist who investigates a given field is not necessarily the best person to explain his discovery to the non-specialist. If he is a good research-worker his whole heart should be in the job and it would be a distraction to have to consider the linguistic and psychological problems of communication. So we introduce a "communicator" whose function is first to understand the original report of the scientist, couched in the latter's own jargon and often well-nigh incomprehensible style, and second to translate this into a form adapted to the needs of the reader.

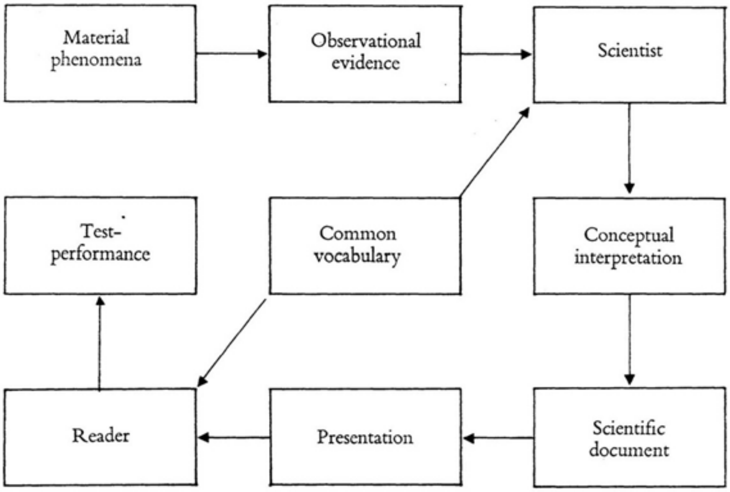


FIGURE 2. Documentary communication.

Now this situation provides us with four patterns of information to which we shall give distinctive names since they are fundamental in the whole development:

I The pattern of information in the familiar subject-matter, called the *Corpus*.
II The pattern of information in the novel predicate-matter, called the *Scriptum*.
III The pattern of information in the communicator's document, called the *Modus*.
IV The pattern of information in the test-results, called the *Modulus*.

The modular calculus is a system of logical variables and rules of operation by which any given pattern of information can be expressed in symbolic form. The pattern thus constitutes a map. The structure of the variables in such a map will be called a map-function. We thus have four types of function, the C-function (Corpus), the S-function (Scriptum), the D-function (Modus) and the L-function (Modulus).

Before considering the variables themselves and the matrices in which their meanings are expanded, we need to give some indication of the way in which these concepts work together to throw light on communication. Each of our four patterns may be envisaged as occupying a screen. From each screen to the next is a kind of optical projection process. But this has to pass through a kind of medium in which the projection pattern undergoes, or may undergo, various modifications. These may be (i) losses, (ii) distortions, (iii) gains. Thus, in general, no two of the patterns are identical, but they will have, or should have, a good deal in common. In addition to the map-functions we therefore have a series of *projection-functions,* representing the action of the medium in modifying the map-functions.

We have already said that the Corpus represents the body of material information with which the reader is assumed to be familiar and on which the communication of the novel information is to be based. An epistemological point arises here. This information is derived from previous experience which the reader is assumed to have had. It is this experience which provides our first projection-function. The directly experienced field of material objects is called *Z*. It is non-linguistic and cannot be incorporated in a document. The variables in which it is expressed in the document are called "types" in the modular calculus. The variables of *Z* are called "sub-typical" variables. The process by which sub-typical variables give rise to types is deemed to be the concern of epistemology. The field of *epistemic communication* is a homogeneous *documentary* field. Thus the first projection-function by which *Z* gives rise to *C* is the business of epistemology and is outside our present province. The modular calculus provides formal definitions of its types in terms of the symbols for sub-typical variables but is not concerned with determining the nature of the latter.

In most documents the Corpus and the Scriptum are rolled into one. The familiar subject-matter and the unfamiliar predicate-matter are intertwined. This is almost a necessity of language, which requires both nouns and verbs in its sentences, and this probably has a good psychological reason. But for our purposes the two must be sorted out and it is part of the business of epistemic analysis to separate the information of the Corpus from the information of the Scriptum. This is not just a matter of separating nouns from verbs, however. It is rather a matter of deciding which aspects of the document are to be assumed familiar and which aspects unfamiliar. *For in any document the familiar concepts provide the material out of which the unfamiliar concepts are constructed.* A document containing nothing but unfamiliar concepts would be incomprehensible. A document containing nothing but familiar concepts would be banal. Communication depends upon the skilful blending of the two. (This epistemic principle is somewhat analogous to the concept of *redundancy* in information theory.)

Thus the projection-function from *C* to *S* represents the process by which the predicate-matter of the Scriptum is constructed out of the subject-matter of the Corpus. The Corpus expressed familiar experience of material things. Latent in that experience are certain unnoticed potentialities out of which new knowledge can be generated. These are made explicit in the Scriptum. In general the actual process of generation demands experimentation. This determines which of the potentialities are in fact realizable. But it does not of itself make them comprehensible. The scientist himself is not satisfied with discovering a law by experiment. He wants to show how the law must have been predetermined in the properties of his materials. The experiment concerns the truth of the law, not its explanation.

It is at this point that the imagination of the scientist injects into the familiar knowledge of materials certain conceptual elements or “intervening variables” by which to relate what is familiar to what is novel. Thus the $C \rightarrow S$ projection function *adds* certain elements to the pattern of information in the Corpus in order to give a coherent Scriptum. The Scriptum is by this much richer than the Corpus.

It should now be evident that in a scientific document in which, as is often the case, there is a description of instruments and of the evidence they yield, together with a theoretical interpretation of that evidence, the reader is presented with concepts of several different orders. The mind of the scientist moves about easily among his own concepts whatever their type, but the reader, to whom they are unfamiliar, may well get lost among different levels of abstraction, particularly if there is nothing in the typography or design of the document to differentiate the levels.

It is thus the duty of the communicator to anticipate these difficulties by making these differentiations himself and by making them explicit in his design of the document. The primary function of the theory of semantic matrices is to provide him with a technique of analysing concepts. The principles underlying the translation of these insights into documentary factors and forms are a separate issue to be dealt with in another publication. The immediate need is to provide a constructive theory of semantic matrices.

CONSTRUCTIVE THEORY

In order to keep the development of semantic matrices from spreading out uncontrolledly, we confine ourselves to four carefully defined types. These have some affinity with logical categories but are designed as semantic elements referring to different “modes of diversification” to use Whitehead's term (4). Given any typical physical environment, *Z,* e.g., a river in which all kinds of events are taking place, we can divide it up in four principal modes: (i) accord ing to the material objects (e.g., pebbles, water, weeds, fish) constituting *Z;* (ii) according to the distributions of various properties (e.g., temperature, colour, acidity, hardness) characterizing *Z;* (iii) according to the functional activities of the various objects one upon another (e.g., erosion, consumption, transport) maintaining the dynamic character of *Z;* (iv) according to the connexions and mutual dispositions of the various objects by virtue of which *Z* forms a connected whole.

The above four types of element are called, *X, K, F,* and *R* respectively. Although these bear some resemblance to logical individuals, classes, predicates, and relations, respectively, they are not to be identified with these. Anyone who has attempted to work simultaneously on the semantic level of physical reference and the logical level of formal inference will know of the pitfalls which logic offers. Roughly speaking logic has become a game for logicians free from the trammels of semantic reference. In semantic analysis we must feel equally free to define our categories in accordance with the demands of physical meanings. Admittedly we must not fall foul of any of the truths of logic, but logic is a system of conventions as well as of truths, and there is nothing mandatory about conventions. It is the conventions with regard to logical "types" which present the semanticist with such difficulties. The attempts (5) of Russell, Zermelo, Skolem, Quine, von Neumann, and Bernays to formulate workable systems of types are, perhaps, to be regarded as experiments in system-building rather than as studies in logical necessity, even though inspired by the necessity to eliminate contradictions. Physical reasoning finds an occasional toleration of contradictions not merely unescapable but even fruitful.

Various graphical devices have been used for representing the relations of inclusion, overlapping, etc., which occur in logical calculi, e.g., the circles of Euler and the squares of Gonseth (6). Their function is to provide intuitive visual evidence for relations otherwise expressed purely symbolically. One curious feature of these devices is their use of *areas,* which are two-dimensional, to represent sets, which are one-dimensional. A linear symbolism will be adopted here for these representations thus setting the second dimension free for the representation of differences between the sets themselves.

The importance of developing an adequate system of representation for these relations between sets is that it is these relations which give rise to equations, and equations are the ultimate and most precise expression of scientific concepts. If we had a completely unified science we can suppose the total range of its variables to constitute a multidimensional manifold or *N*-space. This can be sliced up in various ways into 2-spaces. To express relations involving more than two variables we can arrange two or more of these 2-spaces side by side with suitable connexions. In this way a mapping of concepts from the *N*-space on to a two-dimensional graphic framework can be achieved. For very simple relations a single semantic matrix may suffice. For more complex concepts we shall, in general, need an array of matrices. Thus a semantic matrix is a 2-space in which variables are displayed so as to indicate scientific assertions. We now have to establish conventions for these displays.

CONVENTIONS FOR SEMANTIC MATRICES

We distinguish between (i) *G*-matrices (i.e., general) which may be of any existing forms and for which general conventions either do not exist or have no systematic relations to one another, and (ii) *Y*-matrices (i.e., types) which represent one or other of the four "modular types," *X, K, F, R*. It is these latter that concern us here.

It has been shown (1) that physical concepts of a concrete character can be mapped on to a framework having *X, K, F,* and *R* as coordinates. Given such a mapping we first construct matrices for all the types identified in the analysis. These matrices display the physical meanings embodied in the types, i.e., their *entailments*. But the document is more than the sum of its separate type-meanings. Each type is "syntactically oriented" towards one or more other types by important relations. These syntactic orientations are determined by inspection and each is then represented by an appropriate matrix. Assuming that the document possesses a certain semantic unity, we should next search for relations by which the succession of *orientations* can be seen to cohere into a logical sequence. These *coherences* in turn will be expressed in one or more matrices.

Thus the full analysis of the document will be represented by three sets of matrices: (i) entailment-matrices, (ii) orientation-matrices, (iii) coherence-matrices. These three roughly correspond with the semantic structures of words, or sentences, and of paragraphs, but the analysis may not adhere at all closely to these linguistic divisions because idiom and style may obscure underlying structure.

For the sake of exposition we will posit a fictitious *encyclopaedic store* in which all the variables known to science are listed, and all the values of these variables, qualitative or quantitative, are identified and codified. The whole store, though large, is taken to be finite and discrete. Each variable constitutes a *taxonomic range,* i.e., a set of discrete entities sharing some common character.

In the modular calculus (1) we have distinguished between two kinds of analysis of knowledge, viz., *epistemological analysis,* in which philosophical considerations predominate, and *epistemic analysis* which is concerned with scien tific meanings and the problems of communicating them. Our four modular types *X, K, F, R* are epistemic categories in whose definition certain experimental data are assumed. It is left for epistemologists to determine the status and provenance of these data. The data assumed are as follows:

1.	Environments	En
2.	Locations	Lo
3.	Events	ξ
4.	Structural factors	(hikτ)
5.	Connexity	ρ

The predictable recurrence of an event ξ_c, having a particular character, gives us a property φ_c. A location Lo is the volume of space occupied by a material object in its various positions through space-time. We define objects *X* in terms of locations and properties, i.e., as a set of properties associated with a given location. The *K*-type (which corresponds with the *class* in logic) is defined as the distribution of a given property over a set of locations. Briefly,

$$X = \text{Lo}\Sigma\phi \quad \text{and} \quad K = \phi\Sigma\, \text{Lo}$$

The *R*-type is the expression of a connexity ρ between two structural factors. What this means is that we do not define relations as holding between objects as such since two objects may present an indefinite number of relations. But every object, by reason of its position in the environment, its orientation, momentum, etc. (any aspect of which is called a "structural factor"), exhibits mutual dependence or connexion with other objects. A relation *R* is defined as any one such connexion. (This is not an exhaustive definition. Relations of higher degree are defined separately.)

Functions *F* express the behavioral dominance of one object over other objects. A river erodes a river-bed. A fish eats an insect. A bird builds a nest. In all such cases we have an event ξ initiated in one location producing changes in other locations. The causal connexion ρ is taken as given. Thus *R* and *F* are briefly defined as follows

$$R = (\text{hik}\ \tau)_1\, \rho\ (\text{hik}\ \tau)_2$$

$$F = \text{Lo}_1\, \xi\, \rho\ (\Delta \text{Lo}\ _{2,3 \ldots n})$$[3]

The changes are themselves events. Thus a function may be regarded as one event which *dominates* a set of further events. We can therefore reformulate the definition:

$$F = \text{Lo}_1\, \xi_1\, \rho\ (\Sigma\xi_{2 \ldots n}\ \text{Lo}_{2 \ldots n})$$

In order to achieve the maximum syntactic power from the matrix-arrangement we adopt a standard convention. Every matrix is a rectangular array of

[3] Δ stands for any change.

positions. Each position has two coordinates. A semantic element in the matrix is represented by a small circle occupying a particular position. In general only a fraction of the total available positions are occupied. Each circle is numbered for identification. Its meaning is determined by its two coordinates. These are defined, in principle, by reference to the encyclopaedic store, but in practice by *ad hoc* descriptions (until the world decides to establish such a store). Lines can be drawn connecting pairs or sequences of elements.

Since the coordinates represent variables in the store, i.e., taxonomic ranges, every element represents the intersection of two such ranges. Thus an element is a pair of values of two variables associated together. We now have to examine the grammar of this convention.

Traditional logic adopted the subject-predicate form for the structure of a proposition, "*S* is *P*." There has always been something both compelling and unsatisfactory about this. The confusion can be dispelled if we distinguish clearly between the purposes of logic and the purposes of communication. Communication which gives real information says something new. But it must say it in a vocabulary which is familiar to the recipient, otherwise no communication occurs. Thus looking at a sentence analytically he sees nothing unfamiliar in it. But apprehending the sentence as a complete structure he gets information. For example "The meteor contains nickel" consists of familiar words. Their juxtaposition in that order yields information.

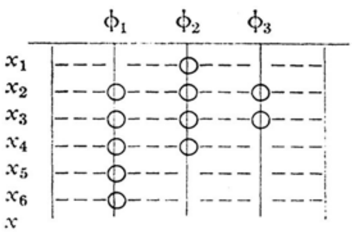
The formula "*S* is *P*" misleads us completely on this. To call "the meteor" the "subject" and "contains nickel" the "predicate" is purely arbitrary. The subject is the set of words *the, meteor, contains, nickel* The predicate is the whole sentence. Thus the predicate is the information yielded by the arrangement and separate meanings of the words. Thus the predicate is always *implicit*. Hence the actual communication of the predicate depends as much on the recipient as on the communicator. Unless the recipient relates the given words together to yield the information in his own mind the communication fails. We may say that the communicator *presents* the subject and the recipient *constructs* the predicate.

Having grasped the predicate as a unified meaning, i.e., a concept, the recipient may anchor it by giving it a name, or a formula or some other label, e.g., "nickeliferous meteor." This then becomes a new term in the vocabulary. It may then, in a later communication, play the role of a subject-term, e.g., "nickeliferous meteors are found in outer space." This process by which implicit predicates are translated into explicit subject-terms represents the consolidation of science. Every new discovery is ushered in linguistically as a predicate. It is then converted into a subject-term to play its part in the expression of further discoveries.

This view of the subject-predicate relation underlies the grammar of semantic matrices. Every element in a matrix is a meeting of two variables. The latter constitute the subject. Their association constitutes the predicate. Then in the same matrix a number of other elements are presented, each representing a two-term predicate. Each such element, by being labelled, becomes a potential subject-term. The association of all the elements in the matrix constitutes a single complex predicate. The structure of this predicate is determined by their mutual dispositions in accordance with the conventions of the matrix. The matrix as a whole can then be given a label. It can, in turn, be associated with other matrices to yield a still more complex predicate. Whether this can be taken in as a whole or built up from the partial associations of pairs of matrices will depend partly on the insight of the recipient and partly on the character of the meaning to be communicated.

Reverting now to the one-dimensional representation of a class we find that this immediately yields a very powerful method for expressing class relations. And since every matrix can be regarded as a system of intersections of ranges of variables, i.e., classes, the method gives a clear interpretation of all the spatial relations in the matrix. Its "geometric grammar" is thus fixed and any implicit spatial assertion can immediately be translated into a proposition in the calculus of classes. This consequence will be called the "inferential potential" of the structure.

The convention is extremely simple. Individuals, here called "locations," are represented by transverse lines across the matrix. Properties (or other qualifying factors) are represented by lines running from top to bottom. Any selection of individuals and of properties can be represented but, in general, they constitute a "universe." Every instance of a property occurring at a location (i.e., every individual member of the class defined by that property) is denoted by a small circle at the point of intersection of the appropriate lines.



Let

$$\alpha = x\ \varphi_1(x)$$
$$\beta = x\ \varphi_2(x)$$
$$\gamma = x\ \varphi_3(x)$$

In this example we see that

x_2, x_3, x_4, x_5, x_6=membership of α
x_1, x_2, x_3, x_4=membership of β
x_2, x_3=membership of γ

Further,

$x_2, x_3, x_4=\alpha\cap\beta$, i.e., product of α and β

$x_1, x_2, x_3, x_4, x_5, x_6=\alpha\cup\beta$, i.e., sum of α and β

Also, since the extension of γ, viz., x_2, x_3, forms part of the extensions of both α and β, we have

$$\gamma\subset\alpha \text{ and } \gamma\subset\beta$$

Further inferences are:

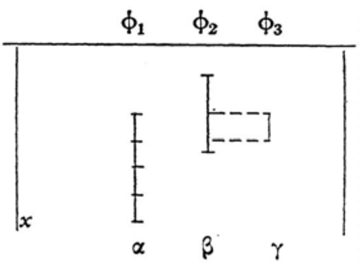
$$(\exists x)\ x\in\alpha\cdot\sim x\in\beta$$

i.e., α has members not shared by β.

$$(\exists x)\ x\in\beta\cdot\sim x\in\alpha$$
$$(\exists x)\ x\in\alpha\cdot x\in\beta\cdot\sim x\in\gamma$$

and so on.

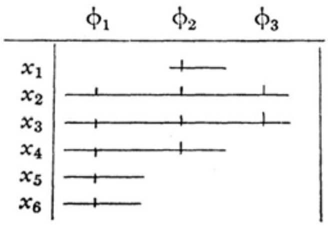
In the above representation we are committed to indicating the actual number of members in each class. If we wish to assert certain class-relations without making the extensions explicit we can replace the columns of circles by straight lines thus:



Here exactly the same relations are exhibited but with no necessity to specify-the number of members in any class. The relations of inclusion, overlap etc. can be shown by parallel projection. As an example $\gamma\subset\beta$ is indicated. The dotted lines show that the projection of γ on β is contained within β.

There is a certain advantage in translating a matrix of discrete elements into one of line-ranges since the latter shows up the space-relations more clearly. But there is a second way in which this can be done. We can draw the lines transversely. A quite new set of relations then emerges. Our attention is switched from classes to the properties of individuals and the latter's resemblances and differences.

Here the range of individuals is necessarily taken to be discrete. The range of properties may or may not be discrete, e.g., it could represent the colours in the spectrum. But there must be points of differentiation whereby one φ is distinguished from the next.



We again have relations of overlap and inclusion, seen by parallel projection. But these give the following interpretations (all of which, of course, are valid only for the universe prescribed by the data):

1. The individual x_1 has a single property φ_2.
2. The individual x_5 has a single property φ_1.
3. The individual x_6 has a single property φ_1.
4. The individual x_4 has two properties, φ_1 and φ_2.
5. The individual x_2 has three properties, φ_1, φ_2, and φ_3.
6. The individual x_3 has three properties, φ_1, φ_2, and φ_3.

Thus

7. x_1 is partly similar to x_2, x_3, x_4.
8. x_2 is distinguishable from x_3 only by its location.
9. x_4 is partly similar to x_2, x_3, x_5, x_6.
10. x_5 is distinguishable from x_6 only by its location.
11. By virtue of these similarities and partial similarities a new class-structure can be discerned.

This last point is a further indication of "inferential potential." We can define new classes thus:

A=the class of individuals having a single property (viz., x_1, x_5, x_6).

B=the class of individuals having two properties (viz., x_4).

C=the class of individuals having all three properties (viz., x_2, x_3).

D=the class of individuals having one common property only, but here we have a *relation* which as formulated is indeterminate.

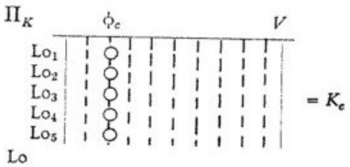
For example x_2 and x_3 both have the same common property with x_1 but a different common property with x_5 and x_6, whilst they have three common properties with each other. This example illustrates clearly the ambiguities of verbal formulation. The matrix representation is entirely unequivocal.

It may here be noted that these matrices are formally identical with the answer-pattern obtained from the administration of tests and examinations in which answers are scored as either right or wrong. The treatment of rows or columns as lines used to be achieved by means of aluminium slats with small cavities to take ball-bearings. These could be transferred to an orthogonal framework of slats by superimposing the latter and inverting. This device is known as a “scalogram” (7). In our example we have considered numbers small enough for individual consideration but the pattern can obviously be regarded also as a statistical distribution.

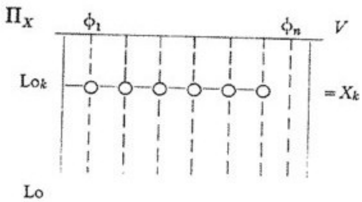
The purpose of the matrix is to exhibit information in a form amenable to the extraction of inferences. The kind of inference to be extracted will depend on the purpose in hand. Since the concept of semantic matrices arose in a context of communications-research in which the problem concerned the comprehensibility of documents we shall naturally stress those inferences which contribute to comprehension. Or putting the matter otherwise *the matrix presents us with an array of possible inferences any of which may be selected as defining what we are going to mean by “comprehension.”* In other words the matrix contains implicitly all the relations needed for the construction of a test of comprehension. Whatever arbitrariness there may be in the selection of relations for the test it will be a manifest arbitrariness which can be defined in relation to a population of potential test items. The selection can thus be made to follow the principle of what Egon Brunswik (8) called “representative design.” Given a total population of inferences which collectively determine “full comprehension” we can suppose this population to be stratified into ranges of inference of different types. We then see to it that each range is sampled in the test by a number of items proportional to the frequency of inferences in the range. The possibility of weighting in accordance with some criterion of relative importance must be allowed for, since there may be key-inferences whose effect is paramount though they are few in number.

One caution is needed here. Although testing of this kind has many parallels with psychometric testing there are fundamental differences which render any analogies between the two dangerous. In psychometrics the object is to measure the abilities of individual *persons* and the distribution of abilities in populations. In epistemic communication research we are measuring the comprehensibility of *documents*. Many of the concepts involved are radically different from those of psychometrics. In particular the aim is to ascertain those factors in the design of a document which make for difficulty and for comprehension respectively. We can then hope to redesign the document so as to maximize the latter and minimize the former factors. The psychometricians do not hope to redesign their human subjects.

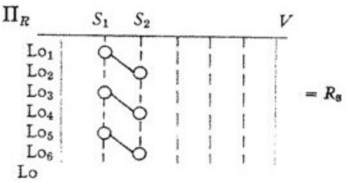
THE FOUR TYPES OF Y-MATRIX



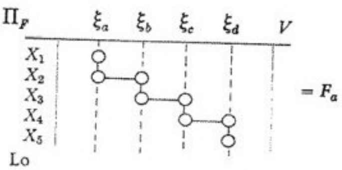
Class-matrix K
A property φ_c is distributed over a number of locations.



Object-matrix X
A number of properties φ_{1-n} are associated at a single location Lo_k.



Relation-matrix R
A specific connexion exists between pairs of structural factors, S_1, S_2 associated with pairs of localities, Lo_{k1}, Lo_{k2}.



Function-matrix F
A dominant event ξ_a at X_1 involving X_2 initiates a sequence of other events ξ_b, ξ_c, ξ_d ....

In analysing the document we aim at expressing all its primary factual content in terms of these four matrix-types, singly or in combination. In other words, every datum is interpreted as the occurrence of a variable *V* at a location Lo or a combination of such occurrences.

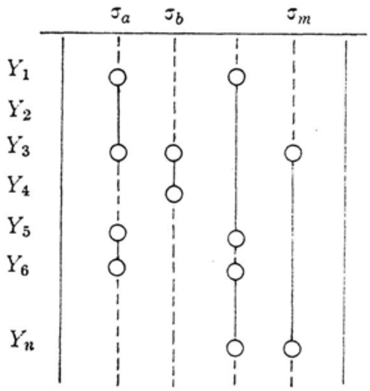
But the document, in general, expresses more than a sum of factual data. We next have to look at the connexions between these data. The varieties of connexion are very numerous. Our programme includes their eventual reduction to expressions in terms of the same four types *X, K, F, R* but of "higher degree" and involving an analytical policy in respect of the use of language. Pending the completion of this programme we rely on a descriptive formulation of the semantic and syntactic connexions between primary matrices and give these "syntactic orientations" a verbal formulation, together with an appropriate symbol.

Regarding the document as initially a sequence of factual concepts expressible as modular types $Y_1, Y_2, \cdots, Y_n$ we note that the sequences of these concepts is partly determined by semantic connexions and partly by idiosyncrasy, style, expository needs, etc. Thus the juxtaposition of two concepts does not necessitate the recognition of any important connexion between them. Connexions may exist between concepts which are widely separated in the document. Thus we now need a means of expression by which *any* important connexion may be revealed. For this purpose we construct a matrix of syntactic orientations. We here determine by conceptual inspection which of these orientations is important and express their logical or other sense in a suitable verbal formulation. Obviously many of the σ's will be left blank.

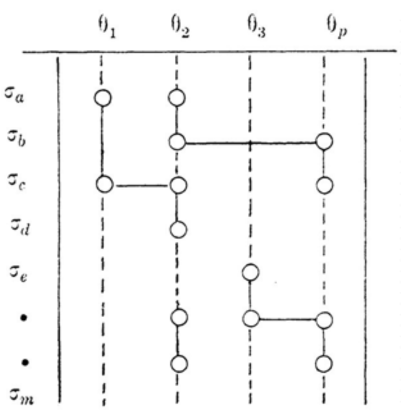
	Y_1	Y_2	Y_3	$\cdots$	Y_n
Y_1	·	σ_{12}	σ_{13}	$\cdots$	σ_{1n}
Y_2	σ_{21}	·	σ_{23}	$\cdots$	σ_{2n}
Y_3	σ_{31}	σ_{32}	·	$\cdots$	σ_{3n}
·					
·					
·					
Y_n	σ_{n1}	σ_{n2}	σ_{n3}	$\cdots$	

Certain types of syntactic orientation will be found to recur. We can ideally envisage a complete system of such connexions, of which a small sample is manifested in any one document. If we symbolize these type-connexions by σ_a, σ_b, σ_c, etc., these may be regarded as constituting a range of relations whose field consists of the matrices $Y_1, Y_2, \cdots, Y_n$. We can then express the whole syntactic pattern of the document by means of an *R*-matrix. This would appear something like this:

The meaning of this matrix is that an important and definable connexion between concepts Y_1 and Y_2, and again between Y_5 and Y_6 is manifested in the document, another between Y_3 and Y_4, and so on.



How much further we go depends partly on the level of understanding which we deem to be appropriate for the given document and for the recipients in view. Some documents would not lend themselves to further analysis. Some have such an inner coherence that connexions of a high order of abstraction can be discerned and identified. To arrive at these we repeat the foregoing procedure, but this time we seek for connexions between the *σ's*. In other words we ask what considerations of theoretical coherence led the author to connect his concepts in such-and-such a way. Here we may well find three or more *σ's* bound together by some logical, stylistic, or expository tie. The author uses his concepts *functionally* in order to achieve a certain result. Such connexions may be of decisive importance in the design of the document and are thus highly relevant to our research. Indeed in the light of the large number of experiments on "Transfer of Training" (9) over the past 50 years we may well look to these abstract elements as highly important factors in the determination of comprehension. Thus our analysis goes on to some such form as this:



This would mean that σ_c enters into a theoretical coherence, θ_1, with σ_a and another coherence θ_2 with σ_d; that σ_b enters into coherence θ_2 with σ_d and θ_p with σ_c, and so on.

We can even envisage an ideal document in which a single unifying concept bears significant relations to every other concept in the document.

CONCLUSION

These methods for the analysis and display of the concepts in scientific documents are initially designed for the use of research-workers in the field of comprehensibility. We are a long way from understanding the factors of documentary design which make for efficient communication, though we have some clues. In such a field as this one of the prime requisites is to secure the conditions for the *replication* of experiments. One of these conditions is a standard methodology and terminology. In semantic matrices and in the modular calculus a beginning has been made towards satisfying this condition. A further requisite is a standard system of non-parametric statistics for the analysis of experimental results. Here the recent work of Siegel (10) is proving an invaluable guide. For experimental work on the psychology of visual displays see ref. 11 on "Planned Seeing."

In view of the frightening discrepancy between the advanced state of knowledge of a comparative handful of scientists and the low level of understanding not only of the population in general but particularly of administrators and statesmen, and even of scientists outside their own special fields it would seem that research on the comprehensibility of scientific documents is a matter of high urgency. Documentary communication is not, of course, the only medium, nor today is it the most popular. But it is important to appreciate that all the other media, graphic, filmic, radio, television, etc., all start from a script, i.e., a document. Unless the concepts and intentions of this document are clear, the translation is vitiated from the outset. Thus documentary research is fundamental to all studies in communication.

REFERENCES

1. G.PATRICK MEREDITH, *Epistemic Communication Research,* Modular Calculus, Part I. *Proceedings of the Leeds Philosophical and Literary Society,* 1958.
2. G.PATRICK MEREDITH, Semantics in Relation to Psychology. *Archivum Linguisticum,* Vol. VIII, Fasc. 1, 1956.
3. C.K.OGDEN and I.A.RICHARDS, *The Meaning of Meaning*. 1923.
4. A.N.WHITEHEAD, *An Enquiry Concerning the Principles of Natural Knowledge,* 1919.
5. W.V.O.QUINE, *Mathematical Logic,* 1951.
6. I.M.BOCHENSKI, *Précis de Logique Mathématique,* 1948.
7. L.GUTTMANN, "The Cornell Technique for Scale and Intensity Analysis." *Educational and Psychological Measurements,* Vol. 7, No. 2, 1947.
8. EGON BRUNSWIK, *Perception and the Representative Design of Psychological Experiments,* 1956.
9. G.PATRICK MEREDITH, The Transfer of Training. *Occupational Psychology,* April 1941.
10. SIDNEY SIEGEL: *Non-parametric Statistics,* 1956.
11. SIR F.BARTLETT and N.H.MACKWORTH, *Planned Seeing,* H.M.S.O., 1950.

Interlingual Communication in the Sciences

JOSHUA WHATMOUGH

The writer of this paper is not a scientist, but a linguist who is interested in the sciences and, particularly, in communication.

First, then, science. Unlike all other study and discovery, the sciences have, with hardly any territorial or national exceptions (perhaps Soviet biology), a global acceptance: what chemistry, or any other science, says in English it says also in Japanese, uninfluenced by any personal, economic, political, national, religious, or other “ideology” or “myth” (the old word for what is now called ideology). In the words of the Report of the National Science Foundation (October 1957):

Science is not contained within national boundaries. The concepts, objectives, and methods of scholarship and scientific research, if not its language, are common to all nations. It is one of the few areas in which there is international understanding. Hence, better communication in this area has possibilities of assisting toward better understanding elsewhere.

Science is, in fact, in a strategic position to attack problems of global communication. Yet the primary and crucial attack is still to be made; and the time seems to be ripe for making it, now that we have swift, sure, and worldwide systems of intercommunication from any one part of the globe to any other or others that have modern transmitting and receiving stations.

The traffic jam that has developed in documentation is the consequence of trying to do first what should come last of all—abstracting, indexing, cataloguing, storage, retrieval and the like—and this in a variety of languages, over and over again, sometimes even in one and the same language. The result is that the entire structure has become superhumanly if not supermechanically top-heavy and overloaded, and is about to collapse of its own weight. The primary research necessary to create a truly interlingual (so-called “international”) means of communication is still to be undertaken: but it is the first step, and the problem of finding such a means of communication that will be

JOSHUA WHATMOUGH Harvard University, Cambridge, Massachusetts.

satisfactory and also supralinguistic (by which I mean as far superior to language as language is to the subhuman communication of the birds, beasts, and insects) calls aloud for solution.

We hear much, far too much from educationists, who repeat their complaint parrot-like one after another *ad nauseam,* of the dangers of specialization. But except for the rare few who have a gift for synthesis, specialization is necessary and inevitable: the specialist certainly has his place, and an important place. The contrast is brought into most vivid relief when a specialist undertakes to criticize a work of broad scope. During historic times there has been no fundamental change in human physiological and psychological capacities, or in human aptitudes produced by biological adaptation. What man has done in the twentieth century of the present era was inherent in man in the fifth century of the preceding era. What *has* changed has been the extension of his "memory" through the means of making and preserving human records, all the way from simple pictographic devices to the most recent and elaborate machinery for recording, storing, coding, and retrieving information. This is the problem of modern man: his brain remains what it has been all along, but the amount of data which he is feeding into it, with its limited input channels, has become vastly larger, causing what is essentially a breakdown of the system, freely recognized as such in individuals. Hitherto this modern problem has been met, but not solved, by the device of specialization, since so few possess the necessary faculty of synthesis, and (more important) since progress in any field depends on specialization. Who, nowadays, can take all knowledge for his province?

But it is the modern extension, in breadth and in depth, of human knowledge, more than specialization as such, that is dangerous, unless it can be overcome by the diffusion of information and by adequate means of intercommunication at a global, not merely national, level. This, so far as I see, cannot be done by merely developing existing methods of communication and documentation, but only by transforming the medium itself, the "code;" for then we shall attack the causes of the trouble, instead of its consequences; and instead, by so doing, of aggravating the present situation which we must seek rather to combat. For the work potential of the human organism is more or less fixed: the deficiency is in the "mechanics" of communication. Abstracting, as now conducted, has become an endless Penelope's web, a taking apart by the less able of what other, and better, men have put together. Much more is required than documentation, which is inadequate, and sometimes quite inferior. A cataloguer in the Harvard College Library, supposing Charles Galton Darwin's *The Next Million Years* to be concerned with divination, actually classified it as Folklore, and I had to point out that the book is an attempt at scientific prediction by a sober biologist! (Incidentally it may be observed that one of the worst muddles in documentation is produced by the attempts, made in large library catalogues, to give, under the guise of a subject index, a bibliography. Those who do not know how to get at the bibliography of a subject ought not to be allowed inside a large library at all.)

In short, the primary research in interlingual communication, in the sciences (and in other fields), should be to reduce the medium of communications to a single form first, before indexing, abstracting, cataloguing, and documentation are attempted—these are the final steps if confusion and duplication are to be avoided.

Next, communication. Communication is a relation between persons, provided that they share one and the same means of communication, and this in its turn is essentially a matter of convention. In the ancient middle east cuneiform writing performed this purpose and the "interpreter" (dragoman) was a rarity; in Chinese the literate now use a single form of writing in the same way, no matter what form (dialect) of the language they speak; and in modern times Bishop Wilkins (1668) and Neurath (1937) have proposed this very solution of Babel, that is, a symbolism once removed from linguistic symbolism.

Linguistic symbolism proper has never succeeded as a means of worldwide communication, even partially. The common Greek of post-Hellenistic times was heard from Spain to the Punjab; imperial Latin from Britain to Asia Minor, and its successor, the written Latin of the church has circled the globe; Aramaic, as a written language of clerks and traders, from Armenia to Egypt and Persia; Arabic, as the language of Moslems, has been in use from Spain to Indonesia. Of these Latin and Arabic have been the languages of the learned, of scholars, and of scientists; but there is no turning back, despite modern attempts at the revival of Latin (*Nuntius Latinus Internationalis* and others). However effectively all these symbolisms may have served the *Oecumene* or inhabited world in the past, even of international scope, they can no longer do so, and certainly not Arabic, the only actual survivor in normal use. The British Association for the Advancement of Science explored the possibilities of Latin in 1921, and decided against it.

Anything that involves the use of a code book, whether based on numbers or words (e.g., simplified English) or an artificial language constructed for interlingual communication involves translation, and therefore is subject to all the delays, expense, inadequacy, and difficulties that *all* translation involves.

The discovery of babel came with the expansion of Europe, the age of voyageurs, missionaries, and traders; the invention of printing, and, in modern times, of the steamship, the telegraph, submarine cable, of the airplane, and now of radio, which have led to a movement for an "international" or "uni versal" language. To the Churchmen of the fifteenth and following centuries there have succeeded the philosophers of the eighteenth, who sought a rational system of signs to convey meaning, the anthropologist, and the scientist of the nineteenth and twentieth centuries.

Attempts to achieve the desired end have been based (1) a priori, on first principles—seventeen such attempts were made between 1629 (Descartes) and 1902 (Dietrich); (2) on the use of existing materials out of which to create an artificial language for worldwide use—over thirty of these are known from the "langue nouvelle" (1765, Faiguet) to Interlingua (1951) of the Internation Auxiliary Language Association; (3) on the revival of a language no longer spoken, usually Latin; (4) on the advocacy of a national language, with modification (e.g., basic English) or without it, i.e., depending on the historic spread of a language that normally accompanies the spread of a common civilization or culture; and in fact already a large number of words, especially of scientific terminology, are available and in actual use; (5) polyglot dictionaries.

In the past it has been the course of events, not the arguments of politicians, philosophers, or philologists, amateur or expert, that have made "world languages;" now the time has come for the interests of scientists, technicians, and businessmen to be served.

The difficulties of translation are notorious. Woodrow Wilson spoke of it as the compound fracture of an idea; perhaps he was thinking of the Chinese version, turned back into English as "invisible idiot," for the original proverb "out of sight, out of mind." The costly devices, or the time-consuming use of interpreters, at the League of Nations and, since then, in meetings of the United Nations, or at international congresses, are equally well known—and disliked. The cost of translations, or even of abstracts in translation, is staggering; there is great delay in making them available, and frequently they are not trustworthy. But the actual trend, the course of events, is and, for at least fifty decades, has been such as to make some form of interlingual communication daily more urgent.

Technically this ought to be a simple matter with modern methods of communication, which are quick and certain. Theoretically also it ought to be simple, for it is an axiom that the use of a common means of communication accompanies ease of communication. Yet linguistically serious barriers exist in the number of languages in use, and in the difficulties to be overcome, and the effort expended, in mastering more than two or three of them. This is true of Europe and of the Americas alone; the difficulties are merely multiplied if the continent of Asia also is considered, as of course it must be. The state of affairs is worse than paradoxical—it is exasperating, frustrating, and vexatious, a hindrance to the spread of scientific knowledge alone, all other considerations apart. Even proper names are not exempt—Pfatten (in the old Austrian Tyrol) is now Vádena, and what used to be called Agram is now called Zagreb.

As for artificial languages, it must be conceded that they have failed. Out of thirty such that have been proposed, beginning with Esperanto, only six contenders are left in the field, and their supporters not only cannot agree which to accept but they even attack one another. Since 1889 over 700 periodicals devoted to the cause of Esperanto alone have been started; of these none lasted more than a few years, only fifteen survived in 1939, and today the number is even less. All artificial languages so far have been biased in favor of European ways of saying things, but there are other ways, more than one, of saying the same thing, and Asiatics and Orientals are not favorably impressed by what purports to be international and then turns out to be entirely Occidental and European.

A language is a code in which messages are transmitted. In other words it is a system of like or partly like recurrent features, together with their corresponding meanings, that is, it is a systematic symbolism which can be transformed into other systems, e.g., electrical (as in the telegraph or telephone) or electronic (as in the machines which have been built to reproduce "synthetic" speech, e.g., the Edinburgh PAT (parametric artificial talking device) built by W.Lawrence for the English Ministry of Supply); as a process, it generates a controlled sequence of symbols from a finite set, with an indefinite population of combinations and permutations (upon which variations of meaning depend). "Language" in the abstract is not a part of nature, linguistic events are; and also those events which the linguistic events (utterances), at first remove, symbolize. The "language" that anyone knows is, physically speaking, at the utmost the residues of traces of his lifetime of experience left as a series of routings and junctions in the nervous pathways of his brain, to form a statistical storehouse of memory, a cerebral assembly of patterns and of links between them, which provide for the controls and impulses that govern his acts of speech each time he "opens his mouth." The mystery of all this lies in the matching with the actualities (meanings) which they symbolize, and with one another, of (1) the speech acts, the phenomena with which linguists deal; (2) the structual forms which these acts may variously take in different languages, sometimes with almost complete agreement (e.g., the first person singular pronoun, "I"), usually with far less, or none at all; (3) the corresponding written or printed marks on paper; or (4) any form of systematic symbolization which may be substituted for these last, and here there is almost unlimited scope for invention and improvement, most of all at an interlingual level. Symbolism (not to be confused with semiology) intervenes not only in protocol or basic statements, but also in subsequent, i.e., logically derived statements, over and above primary situations, and this, as we shall see, introduces further difficulties.

Linguistic change, and interlingual equivalence, are both manifestations not of independent and isolated phenomena in series, but parts of a system or systems. Historical linguistic phenomena, that is to say, are parts of a given synchronic structure or status, and notwithstanding the possibility, and fact, of occasional "holes" in the pattern, there is such a thing as the "economy" of linguistic change. This has been revealed most clearly by the shift of interest from exclusively qualitative phenomena to the increasing study of quantitative linguistic phenomena.

Here I venture to refer the reader to some paragraphs in my book *Language* (London and New York, 1956, pp. 251–257; or in the paper cover edition, New American Library, 1957, pp. 224–229) on the general topics of the relation between language and the sciences, and to urge the reader to ponder them before preceding with the rest of this paper.

Since there is only one human species, it is obvious that all its members may well and ought to have a single means of communication. But all past experience shows that is not likely to be merely linguistic, which inevitably involves either prestige or translation. Human translation is riddled with shortcomings; and it has been asserted that "the simple basic fact that the meaning of every linguistic expression depends only upon incomplete, uncertain, and temporally shifting convention within a community of language has as its necessary consequence that equivalent reproduction of a text in a foreign language is impossible." Gross blunders may have, and have had, serious consequences, even though the translators were believed to know what they were doing. The mathematical theory of communication furnishes proof of what was already well known, the degradation of any translated message.

In the meantime, that is, until a supralinguistic procedure can be devised and perfected, it therefore seems best to depend on the hope that mechanical translation may be achieved. Such translation has limited aims; it does not aim at perfection, but practical utility. Of the prospects of success, and of the uses of mechanical translation, I must leave Professor Oettinger to speak. He is an acknowledged expert and an active worker in the field. Here I insist only that documentation, abstracting, retrieval, and the like, should be performed on one language only, after foreign sources have been converted into it: this in the interests of economy both of effort and of money, as well as of simplicity and, in the long run, of effectiveness, granted that word-for-word translation usually needs exegesis—witness the early translations of the Bible into Gothic, Armenian, Old Church Slavonic, and other languages, all of which were done on a word-for-word basis.

To put the matter another way, linguistic analysis above the levels of phonology and morphology is still largely to be done: such things as word order, syntax, and even meaning are largely unexplored, at least in such a way as to be practically useful in mechanical translation. Attempts made by one or two logicians to cope with the problems presented have not so far given much promise of filling the need; indeed one of the foremost of them, Chomsky, explicitly declines to consider the problem of meaning at all! More promising are statistical investigations, which are more in accord with the mathematical theory of communication.

Here let me quote from a report on Mathematical Linguistics which I presented to the Eighth International Congress of Linguists held at Oslo in August 1957, and from some of my remarks made in the course of the discussion that followed.

In modern times statistical and mathematical methods applied to language were introduced by George Boole, who in his *Laws of Thought* (London, 1854, p. 245) records how the principle of determinate frequency was applied over a hundred years ago to deciphering Ogam and cuneiform; a number of frequency counts were made during the last century, e.g., by Förstemann, Lanman, and E.V.Arnold, to name no others; quite recently appeal was made to the same principle (of determinate frequency) in helping to decipher Linear B. It was the subject of a Harvard doctoral thesis in Linguistics in 1929 (*HSCP* 40, 1929, 1–95), begun two years earlier. The doctrine was not well received at the time, although the actual objections were far from being reasons (or, if there were any, the reasons for objection were well hidden below the surface). What quickened matters, and brought a mathematical statement, was the enormous progress made by communication theory and information theory during and since the war of 1939–1945, to which, on the linguistic side, Mandelbrot's *Contribution à la Théorie Mathematique des Jeux de Communication* (*Compt. rend. acad. sci., 232* (1951), 1638–1640, 2003–2005; afterwards *Publ. de l'institut de statistique de l'Univ. de Paris, II,* fasc. 1–2, 1953) and Oettinger's recent studies (Harvard dissertation in Applied Mathematics, 1954) have made notable contributions, now being further developed in the seminar in Mathematical Linguistics conducted by Dr. Oettinger, some of the fruits of which it is hoped will be published in the near future.

Then there is the work of Shannon, both his *Mathematical Theory of Communication* (Urbana, Illinois, 1949) and a more recent paper on *Prediction and Entropy of Printed English* (*Bell System Technical Journal, 30* (1951), 50–64), which raise two important questions. In the first place there is the question of the relation of the written language to the spoken. As to this, observe not merely that writing is systematic in its own right, as has been shown by Pul gram and by Uldall independently, but also that even a system of writing that is not phonematic obeys the law of great numbers—this was shown specifically for English "spelling" by Hultzén and Allen at the Chicago meeting of the Linguistic Society of America in December 1955. Then there is the matter of the transformation of one form of energy (and language is an ἐνέργεια as the Greek philosophers recognized, in whatever sense they used the term of language) into another (e.g., electrical), and since Shannon's work brought this problem to the forefront there has been a great deal of misunderstanding. So far I know no one has identified physical entropy with the quantification of "information" (see my specific denial, *Language,* London, 1956, p. 201 note), where not entropy but ectropy (so called "negentropy") is involved, and statements to the contrary, implying such an identification, are false. What *has* been said is that "Information" theory is a useful model of the working of language. It is also false to insinuate that the model, and the theory, infer a purely mechanistic view of language, as if it were all chance and no choice, restrained and directed as the choices are. But attention should be called to recent views of energy as developed by the Danish scientist J.N. Brønsted (died 17 December 1947) in his papers (1936 to 1946) on *Principles and Problems in Energetics* (New York and London, 1955) which deal with the transmission and transformation of energy, equilibrium, and coupling of different systems in ways more refined than the classical treatment of heat and energy. It might well be fruitful to enlist the cooperation of an expert in this field with an open-minded linguist, a neurologist, and a physiologist in order to discover exactly what correlations there are, if any, between language (both spoken and written) and other forms of human activity and consumption of energy. By contrast the work of linguists alone, who must depend upon more elementary knowledge, especially in mathematics and statistics, necessarily leaves something to be desired. Yet there are some praiseworthy, courageous, and unpredjudiced endeavors being made to reduce linguistic phenomena and findings to mathematical formulae, a few of which deserve to be mentioned here, such as the work of Lees and Swadesh in lexicostastics (since about 1952); of Greenberg's quantitative approach to the morphological typology of languages (1954)—producing arithmetic indices that enable the degree of "polysynthesis," "inflexion," or "analytical" method (when word and morphome are identical and frequently monosyllabic) to be compared with precision, both with one another and in various languages; the objective criteria (devised by Annemarie Schlismann) for comparing the styles of authors (Innsbruck, 1955, cf. *Word, 12* (1956), 293–298); correlation methods of comparing transitional ideolects by D.W.Reed and J.L.Spicer (presented to the Linguistic Society of America (*Language, 28* (1952), 348–359); H.H.Paper's formulae for the depiction of symmetry in language (also presented to LSA); Pulgram's formula for prehistoric linguistic expansion (1956); the older works of Zipf, Chrétien, Kroeber, and others. That there is ample scope for the development or correction of these results and, above all, for far greater rigor of method, as well as for new investigations, by as many as feel themselves attracted to or competent in this new and rapidly growing field of Mathematical Linguistics, goes without saying. It has the great virtue of revealing new, unsuspected, and wider horizons, exactly like the application of mathematical logic to biology in recent decades. Improvement of scientific method has always been a matter of science lifting itself up by its own intellectual and linguistic bootstraps; when these are also mathematical (which after all is only the Greek for intellectual), a subject, no matter what prior claims may have been advanced for it as a "science," has at last become truly scientific.

I wish also to call attention to some previous presentations of the relativity of distribution of vocabulary, in particular those of Mlle. M.-L.Dufrénoy (on orientalizing influences in French romance), and of Professor Josephine Miles (on the continuity of poetic discourse in English).

I must, however, point out that nearly all the work done so far, by Yule, Guiraud, Herdan, Zipf, and others is not, everything considered, altogether above the criticism, either from the linguistic side or from the mathematical side, of being somewhat amateurish. For something more rigorous it is necessary to turn to the communication engineers, and their reports as published in the *Bell System Technical Journal,* in the *Transactions of the Professional Group on Information Theory* (Publications of the Institute of Radio Engineers, New York), and of the *Symposia on Communication Theory* and on *Information Theory* (London, 1953, 1956). It is not, indeed, clear that there has yet been a meeting of minds even here. Some of the linguists seem to me to be hidebound by old fashioned notions of language, and most of the communication engineers to be groping in the dark to get some notion of linguistic structure and function, and some others of them to be too much occupied with non-linguistic problems to concern themselves with purely linguistic theory, scientific or not. Moreover, when mathematics is despaired of, as it is by some investigators, it is always possible to resort to symbolic logic, perhaps a necessary preliminary way of formulating problems and solutions, especially in syntax, that may lead through mathematical logic to pure and not merely applied mathematics.

It is, however, quite clear that by means of the methods of statistical description of languages and of applied mathematics, for example as employed in attempts to devise automatic dictionaries and, even, by means of switching theory, automatic composition and translation, closer approach is being made to understanding and formulating the sequential processes, coding systems, and functional symmetries of languages and of language, difficult as problems of context and syntax, idiomatic expression, variants of vocabulary (choice), and semantic alternants (synonomy and homonymy) are. In the widest sense of the term the processes are mathematical, and come within the domain of modern algebras. It is equally clear that there is need of mathematical linguists, that is, workers whose training lies both in modern linguistics and in modern mathematics, subjects not likely to be united in recent times once schooldays are over; and that there are great and important discoveries in store, far more important than anything to which comparing (say) the length of sentences in Marlowe and Shakespeare, is likely to lead.

Most continuous discourse is a matter of the making of routine linguistic decisions according to rules of habit and of choice, no matter what the language. Moreover it is not difficult to formulate these rules: they are to be found, if not in every grammar book, then in every complete description of a language prepared on structural principles. Each sequence in the stream of speech may be seen as a binary choice between that which is accepted (1) and everything else which is not (0): each phoneme, e.g., *s* or not -*s;* or each morphome, *ness* or not -*ness;* or each part of speech, a noun or not a noun; each construct, of phonemes (a phonematic construct, i.e., a morphome) or not; or of morphomes (a morphomatic construct, i.e., a word or phrase) or not, active or not, a sentence (a construct of epilegmata, "words") or not—and so on, to the end of the utterance, whatever its length, a paragraph, a poem, a chapter, an act of a play, a book, an entire play, a whole library, all the libraries in the world. Such a mass of material, however, is unmanageably great: then we take random (i.e., unbiased) samples. For the nature of language is such that a sufficiently large sample will portray the whole, and languages obey laws as stable as any that have been found in human behavior, if not as stable as those of astrophysics (which are not entirely stable). If languages did not so obey, thanks to grammatical rules and lexical conventions, linguistic communication would be impossible. The number of linguistic phenomena in a language is not infinite—phonemes, morphomes, epilegmata, constructions; even the number of orders of arrangement itself, though much larger, still is not infinite: restraints are always imposed in the interest of understanding, of sense, of meaning. A highly convincing mathematical basis, underpinning the metastable equilibrium, historical and contemporary, of language, already found empirically by statistical investigations of linguists and communication engineers, is now actively being developed along lines of investigation that may fairly be called purely theoretical, that is, concerned with the formal and abstract linguistic order (pattern) that exists, and purely for its own sake. This task will not be carried much further, much less completed, without enlisting the interest of the best mathematical minds. The theory of abstract sets is perhaps the most promising point of departure, since in language the one-to-many relation is at least as prominent as the one-to-one. Variation in categories, parts of speech, orders of arrangement, reversed order (as in interrogation in some languages), the presence or absence of infixation, inflexion, compounding, even rhythm and the like, the alternants presented in syntactical rules and a number of other variants that readily come to mind, all make for a considerable number of separate problems, none of them easy, that must be solved before a comprehensive theory can even be attempted. Context, nuance, style present quite terrifying, and still very remote, quests for the future. Alternations, already shown by sampling to rank high in frequency of occurrence, may not be the simplest to solve, practically useful as such solutions would be. Function and "meaning" will have to be differentiated, and especially units or categories that are concerned solely with functional relationships (e.g., prepositions) from those which are concerned with content of meaning, to use for the present terms that are convenient rather than trenchant. The identification of context, and perhaps even more, of word classes (as distinguished from class functions) and patterns of word classes, will call for recursive rules (rules of replication) in the formulation of tokens of sentence types with their recurrent structures. Some preliminary work of a descriptive kind (the analysis of discourse) only serves to show how much still remains to be done. The implications of conventional punctuation, the degree of internal coherence in a text (determined by subject matter), idiomatic expressions (that is, with respect to a given language), redundancy, unusual sentence patterns, partial incoherence or garbling—all these will have to be considered. It is a large program, but a stimulating one that calls for many workers, and may well produce theorems of elegance as well as of conviction; and, at last, a true mathematical theory of discourse.

Actually we may justifiably see, statistically and mathematically, mass linguistic regularity, when apparently some individual linguistic elements follow no regular laws, and without any appeal whatever to metaphysics. Language is regular; individual units of language show variation within this regularity. That is all.

In the meantime, and as a starting point, it may confidently be asserted that the mathematics of communication systems serves as a suitable model, beginning with any linguistic system. The factors involved not only in the linguistic system, but also in speech and in actual speakers, even the physical factors, articulatory and acoustic, in terms of force, intensity, energy, and other physical properties, will all have to be obtained before the time comes for anything like a general theory. This is a program of work for years to come and for generations of workers. Only the merest beginning has yet been made to obtain physical correlates to analytical descriptions, to determine morphome or word boundaries; yet these are among many necessary preliminaries. Any text is a pragmatic affair, for it is a sequence of physical events, spoken or written, produced, transmitted, received, and identified—even sometimes reproduced and retransmitted—by physiological and neurophysiological organs; or it may be handled at certain points electrically or electronically. The preconceived assumption that linguistics, physics, physiology and neurology, force and energy are all completely independent of one another is precisely what has hindered and still hinders progress, most of all progress in linguistics. Actually they appear to be correlated in the most intimate fashion. But the only kind of unitary theory that will be able completely to comprehend this intimate correlation will be mathematical.

The statistical and mathematical methods do comprehend a complete description of the constraints operative upon combinations of the basic elements of linguistic systems, including meaningful utterances, that is utterances permitted, or at least not prohibited, by the system. In theory this may seem an impossible task; in practice, therefore, it is necessary to set, at least for the present, a practical goal, namely that of reasonable approximations. To do so is not unscientific, but rather characteristic of scientific method, and a clear evidence of confidence in linguistics as a discipline that is at last becoming both scientific and mature.

Notoriously syntax is difficult to formulate both accurately and exhaustively. Descriptive structural method has declined the task for thirty years, and still shows little appetite (or should I say aptitude?) for it. But mathematical logic already has made some notable, if only initial, attacks upon syntactical problems—the work largely of Spang-Hanssen and Yngve (on the analysis of gaps), of Bar-Hillel and Chomsky (on sentence analysis), and of Epstein (on grammatical categories)—in each case with results that are stated mathematically or symbolically.

The problem, most broadly stated, is to find the statistical structure of natural linguistic messages, all of which, in all languages, are decomposable into *epilegmata,* language being a sequence of discrete entities,[1] as de Saussure and others have held. A message may be represented by a continuous carrier (e.g., magnetic tape), but it must be divisible into a set of discrete or symbolic units if it is to be received, indefinitely retransmitted, and understood. If language did

[1] See Whatmough *Language,* 1956, p. 115; *Forms of Discourse,* pp. 43, 116; ἐπίλεγμα a (longer or shorter) *extract* from the stream of speech; cf. ἐπιλέγειν *to repeat* or quote a unit of expression segmented or selected from a whole string, each unit of which is also repetitive in the language, i.e., it has (in the sequence) a clearly marked beginning and end, and also strong internal statistical probabilities, since it has a meaning; ἐπιλέγδην *selectively*.

not exist, and we set about to create it, it would be necessary to string together sets of discrete segments—phonemes, morphomes, words, and so on, sentences, paragraphs, chapters. To deal with the totality of all texts is the unattainable ideal of the structuralists; mathematically and statistically this is neither an ideal nor even a necessity, since a sample text, or sampling of texts, suffices, provided that the sample is long enough to exhibit all the characteristics of a linguistic status. Difficulty arises from individual variations; but, as we have seen, these are covered by mass regularities (macroscopic), however great in practice must be the freedom permitted in a given microscopic problem—and "every case is a given case when you come to it." Given the range of the number of epilegmata in a sequence of a selected size, and the weighted average of their properties of transmission (e.g., spoken or written), in all permitted combinations, we may determine all the possible messages. These will correspond to actual messages, i.e., agree with the mathematical theory (cf. Mandelbrot's formula of a text),[2] which is confirmed experimentally if we smooth out the steps in a crude plotting of the data.

A language may be either a closed system (e.g., classical Latin) or an open system (e.g. modern English). The status of the former (closed) may be described as a stable equilibrium, of the latter (open) as metastable, i.e., such that an infinitely small change will disturb the equilibrium. The fact of the universal existence of a linguistic status is known statistically.

Statistics distinguishes between finite, hypothetical (i.e., indefinite) and infinite populations (so Yule and Kendall). A language may be said to constitute a hypothetical population—only so can we explain the ability of a speaker to produce and understand certain new sentences and to reject others. That is, there is a finite state grammar, with a finite grammatical apparatus, that generates an indefinite number of sentences.

The system consists of a sequence of binary choices.

Language is a sequence of physical events, controlled by the nervous system which certainly behaves as an *all or nothing* mechanism. The conditions of communication depend upon a correlation of linguistic, physical, and physiological features. "Information" (in the technical sense) is concerned with non-conformation, surprisal, change in pro-presentation, the unexpectedness of the event (i.e., choice). Even "nonsense" is communicable, but at high cost, e.g.,

[2] $P_n=P(n+m)^{-B}$. For this is the frequency of the nth most frequent epilegma:

$$P_n = \frac{P}{n} \quad \text{where} \quad \sum_{n=1}^{R} P_n = P\sum_{n=1}^{R} \frac{1}{n} = 1,$$

provided that m=0, B=1, that is to say, only in a critical case.

William Blake, James Joyce, Dylan Thomas, Gerard Manley Hopkins (all of whom have been regarded as eccentric, if not demented). Their “open” vocabulary has been described (by Mandelbrot) as a linguistic “high temperature;” in somewhat similar manner Yngve speaks of “open” positions within the regular syntactic structure, a sort of cement, this last constituting those aspects of a sentence which have the attributes of a calculus.

Shannon's *Mathematical Theory* (1949) was misunderstood by some early linguistic reviewers (e.g., Fischer-Jorgensen). More recently (1957) Shannon has written as follows: “Language gradually evolves under continued use as an efficient communication code, i.e., human beings tend to adopt an approximation to the ideal encoding of communication theory.” This is what was implied, without a mathematical proof at the time, by my *Theory of Selective Variation* (1941, Harvard examination papers; 1948 Actes VI^{e} Congrès des Linguistes; presented also at the summer meeting of the Linguistic Society of America, Berkeley, California, 1951), viz:

A linguistic status is produced by consistent *selection* which preserves the system from all possible or actual inconsistency of *variation,* thus giving a coherent, homologous, and significant pattern.

It is an elementary observation that transformations of communication are easy, not only from one code to another (as in translation) or by heliograph or semaphore, but also electrically or electronically (telephone, radio, teletype, or radiophoto, e.g., of the *N.Y. Times* to San Francisco during the Republican convention of 1956. These are all transformations of one form of energy (the spoken or written utterance) into another and then back again. Information theory treats a language system as a sequence of physical quanta of energy. Its mathematical formulation is applicable precisely because mathematics is concerned with structure, exactly as any language system is; the mathematical formulation, however, is both more exhaustive and less exhausting than articulate structural methods (as witness the work of Harris and Hockett).

Moreover, using methods analogous to those of combinatorial mathematics, we may apply it at different structural levels, as of phonemes, morphomes, epilegmata, and so forth. The continuous carrier, whatever it is, is segmented into uniform but discrete entities, whatever they are, and it is these which are received, retransmitted, and understood. The evidence so far available indicates that it is the epilegmata which are decoded and encoded one by one, with some slight time delay. No doubt the process would be more efficient if the entire text were coded simultaneously, but this is perhaps not practicable except in radiophotography of short texts. A compromise might be sentence by sentence. Hence arises the need for knowing the structure of sentences. In what follows English sentences have been used; but the procedures are applicable in any language. The text should contain continuous samples long enough to exhibit all the basic characteristics of the language, but not so short as to exclude variations.

It is at once apparent that conditional probabilities are involved. Sentence structure is not haphazard, but subject to syntactic restraints. "Artificial" languages, including the utterances of children who have not acquired complete control of their language, and perhaps also basic English with its limited vocabulary, do not satisfy a general law, but the free-flowing language of schizophrenics is said to do so, a curious phenomenon the explanation of which is still to be found.

Two points are to be noted: (1) the maximization of "information" is completely analogous to the minimization of free energy (i.e., we have to do with ectropy, so-called negentropy or negative entropy); and (2) there are marked transitional probabilities (i.e., the transmission of each quantum restricts the choice of subsequent quanta that may be transmitted).

Coming now to structure, we must ask how this can be described. It is difficult to see how meaning can be used for this purpose, though Mackay's definition of meaning (*Information Theory,* 1956, p. 219) is suggestive, viz:

the selective function of the message on an ensemble of the possible states of the conditional-probability matrix [of behavior patterns] in relative circumstances.

Is the number of meanings limited? And if so, by what? Roget classified the range of English epilegmatic meanings under 1000–1200 headings, Buck the meanings of I.Eu. roots under something over 1300. And if each such unit is regarded as having one (and only one) meaning, then a sequence of three such units (a sentence of three "words") would occur as one of 1.7×10^9 theoretically possible sequences, of which not all, indeed only a fraction, occur in practice, since "nonsense" is excluded. Moreover it is necessary to distinguish between signs and symbols, symbols being open-ended (the number of meanings not being merely one-to-one), and also different symbols with identical tokens (e.g., *bank, walk*) not commonly distinguished except by context. So far, therefore, no means has been found of analyzing structure solely on a semantic basis. More promising is the attempt at analysis on the basis of word classes and functions (see Harvard Papers: Seminar in Mathematical Linguistics).

Theoretically a sentence of 50 words each of which is a token of two symbols would be one of 10^{15} possible sequences, but the actual (i.e., permissible) number is much less, as this sampling showed. It is higher, however, than proposed by Fries, Bar-Hillel, or the Wundheilers, none of whom found any solution to the problem of discovering a correct set of indices.

Sentences show structural ambiguity, and at first glance this calls for a complicated analysis. However, orders of arrangement govern not only sequences of epilegmata, but also sequences of word classes. There seems a remote possibility of classifying tokens uniquely (i.e., into a given class), and this presumably might be accomplished by contextual analysis, using redundancy (i.e., the presence of more clues than are actually necessary) in structure to compensate for ambiguity. Any sequence consists of dyadic combinations of binary-choices and the problem is to describe the degree of selective variation in the structural symbols, i.e., where there is identity of relation but multiple structure.

The constraints are greatest at or near the sentence boundary. Some other observations: if W_2 is class 12 (verb), this fact supplies much information containing the occupant of the position W_1. Class 3 (adjective) has low influence both forward and backward; class 5 (article; “no,” “each,” “some,” “all,” “any”—Russell's logical words, i.e., they indicate structure, and may not be changed without changing it) has an exceptional forward influence. Rather high variation is permitted by classes 1 (substantive) and 4 (adverb). X^2 tests indicate that no two adjacent positions are sufficiently similar to be considered as having been drawn at random from a given population (Shannon obtained the same conclusion in his study of English word order).

What are called “creative acts of the imagination” seem more likely to be variations permitted within the natural laws of language, the intrinsic regularity of which is such that, given sufficient bulk in sufficient variety, no intercepted messages remain forever undeciphered. Since its regular macroscopic properties represent a metastable relationship, which preserves equilibrium for equilibrium's sake (i.e., for reasons connected with intelligibility), notwithstanding the random behavior of its microscopic constituents, the same mathematical methods may be applied as in dealing with the properties of matter in relation to the behavior of its constituents. Large scale studies of sentence structure will in all likelihood reveal (1) logical words, few in number, short in length, frequent in occurrence; (2) object words (“dictionary” words), numerous, variable in length, less frequent in occurrence. The former class, if changed, changes the structure of the sentence, and therefore its meaning; any of the latter may be replaced by the dictionary definition.

How far is it possible to attach quantitative measures to the relationship “language:meaning”? This problem has for the most part been shunned; perhaps the time is coming when it must be attacked more deeply by “information” theory, which so far has barely got beneath the surface.

Symbolic logic, insomuch as it also (like the algebra of abstract sets) deals pre-eminently with structure, may be serviceable in solving some problems:

for example, symbols enter into relations which are reflected in the sequential arrangements of their corresponding tokens; then these relations become symbols in their turn, the features of which may readily be described by logical methods. Syntactical relations (among the tokens) should not be confused with operations, e.g., 2+2 is not a sentence, but a structured name for the integer 4, no matter how stated (Housman, Asquith, Chaplin).

More attention should be paid to punctuation. There was distributed at the Fifth International Congress of Linguists a monograph (printed at Göteborg, 1939) in which patterns of structure indicated by phrasal boundaries were identified in fourteen principal European languages. It was apparent that transition probabilities are involved here also. Hence a sentence may be viewed as a sequence not only of word classes but also of phrasal and clausal classes.

A number of schemes using single syllables, combined with numbers (1 to 9), but demanding the use of a code book, have been devised for multilingual communication, particularly at airfields; none has proved satisfactory, because of the delay involved; attempts to manufacture more or less artificial languages for international use are equally unsuccessful (see *Language,* 1956, pp. 53–58). Some method of supralinguistic symbolism, to which, as we have seen, the sciences are already much inclined, will have to be devised—of this more below. Meanwhile translation is all that can be done, notwithstanding all its imperfections, but it is perhaps the best hope, especially if mechanical translation can be perfected for practical use, as no doubt it can. So far it proceeds practically as word-for-word translation, and this is often roundly condemned. There is, however, something to be said for a word-for-word procedure. In the use of language memory, which depends on input, output, and feedback, plays a large role, even in “creative” writing. It uses chiefly the existing pathways and switching points in the brain; in original composition it creates new ones, which the reader or hearer will decode. All the evidence thus far available tends to show that this will be done, in any particular language word-by-word, not, as might superficially have been supposed, in groups of words, or idea by idea, except possibly in formulaic composition, or in some linguistic types, where the arrangement of words, in accordance with “grammatical rules,” may go into more or less stereotyped order (phrases). New neurological investigation may compel a modification of this statement; but the best research so far conducted views very unfavorably the notion that we receive language, physically speaking, by “ideas” rather than through the words which infer them.

Mandelbrot has shown that linguistic structures, characterized as they are by regularity of pattern, which is governed by probabilities of choice, are adapted to word coding with a minimum of effort, the “economy” of the mechanics of language (glossodynamics, or glossostatics); in his words a message is purposively produced in such a way as to be decoded word by word in the easiest or "least costly" possible fashion.

Difficulties arise because of the difference in pattern in different languages, as well as from differences in meaning in different contexts even in one and the same language.

The adoption of a single language (English, Spanish, Russian, Latin) for global use is not only unlikely in a divided world of many cultures that are still poles apart: it has been attempted, and found wanting. The same is true of "artificial" languages, all the way from Esperanto to Interlingua (and basic English). Yet we understand well enough what the nature of communication is and what the nature of language is. Both are characterized by symbolization, by system or good order, and therefore both lend themselves to coding on a single, universal, and global level. There are grave difficulties; but they are not insuperable—what the human mind has done in creating babel, the human mind *(ex hypothesi)* can cope with.

Lines of attack that suggest themselves are: (1) study not only of the nature of symbols and tokens of symbols but also of the combinations and relations of symbols one to another, even of the relations of such combinations one to another, which then become symbols and tokens of symbols at a higher level in the hierarchy of communication; (2) study of the nature of the cerebral processes that take place when overt meaning (as in public utterance, written or spoken) is suppressed (as in "thought"), which should lead (3) to a more profound understanding of the nature of meaning itself; (4) the invention of means of combining form (symbolization analogous to that which is used in logical syntax) with content (what is commonly meant by meaning).[3]

This last is a challenge to communication engineers. One thinks of switching theory as providing solutions to not dissimilar problems, problems of a far more elementary kind. One imagines (for imagination is needed at this stage) that the final result will no more resemble language than a sewing machine looks like eight fingers and two thumbs. It goes without saying that electronics is the springboard for a leap to the universal "grammar" that I venture to foresee, if I am not quite bold enough to prophesy. I do not believe that it is be

[3] The best statement so far, *for this purpose,* known to me is that of D.M.Mackay (*Information Theory,* ed. Cherry, London, 1956, p. 219), quoted above. The distinction between public (communicable) and private (non-communicable, emotive, a matter of conviction), which together make up that repository which we call cognition or knowledge, has been much debated. There is a brief but good statement in Hogben's recent book *Statistical Theory* (London, 1957) pp. 7–10, which I commend to the reader. He distinguishes also between knowing (as a process) and knowledge, which is a matter partly of recognition, partly of communication. It is, of course, only the latter with which we are now concerned.

yond the ingenuity, courage, and persistence of modern communication theory, or of the builders of those machines, Univac and the rest, that, at present in limited fields, in the future in all fields in which men think, do better than unaided humans.

Once created, the universal "grammar" (I use the word for want of a better) of communication will reduce the labor and cost of documentation, storage, and retrieval, by at least half; and will at least double its accuracy and trust-worthiness.

It has commonly been held that one cannot communicate, at least at any except the lowest levels and in the simplest terms, without communicating in a particular language. This dogma, which is valid to a certain point, lies at the base of all the strife over the language question in Indian, African, and Asian universities. Macaulay held that English and English literature should be taught in the Indian universities the establishment of which he advocated— in those days literature, significantly enough, was not taught at all at Cambridge, or in any self-respecting university. There are those who consider that the use of Maltese in the University of Malta is a brake on the island's progress, and Livingstone held that the survival of native languages (to which Afrikaans may now be added) was a dividing force in Africa. But even language, with all its fine qualities as an instrument of communication, is far from perfect; this unique power among animate creation, is certainly capable of improvement, of sharpening, of finer perception and clearer expression.

Language represents an enormous step forward, above and beyond the means of communication used by sub-human animals. Is it not conceivable that methods over and above language can be achieved? The development of Mathematical Linguistics, of parametric artificial talking devices, of a universal grammar that would truly be infinite in scope—all depending upon further development of switching theory and a better understanding of the nature of meaning, both private and overt, i.e., of the content of expression—this is the real crux, together with better means of relating symbols and tokens, and of these, in their turn, as symbols and tokens of symbols, in a hierarchy of language (here switching theory might surely play a part)—all this suggests that such an achievement is not inconceivable, however far it may be in the future. It is moreover a necessary preliminary step to true interlingual communication, which should be simpler in the sciences than elsewhere; and this a necessary preliminary step to be taken before abstracting and indexing can be anything but chaotic. The recent announcement that computators can develop the capacity to learn, as well as to operate, is definitely encouraging.

If the notion of evolutionary emergence is correct, there is no reason why the present symbolic level should not be vastly improved, if not surpassed.

Human experience is not limited to a mere slice of the universe in the way in which it is to specialized creatures (biologically speaking), each species of which by its very specialization is restricted in what, and how much, may enter its "ambient world." What reason is there for supposing that man has reached the limits of scientific and technological progress in the scope of his peculiarly symbolic activity? In any event all that we call science and technology is systematic symbolization, not merely the concatenation of symbols, but combinations and arrangements of them in such a way as to enhance their symbolic value, a grammar, not just a vocabulary. This may be fertile and productive without limit. A formula is a symbolization devised for a particular symbolic realm so that it becomes part of a great "thinking" mechanism, which, in its turn, imposes what I have called language engineering (see my forthcoming *Lowell Lectures,* 1957)—the next forward step.

An Overall Concept of Scientific Documentation Systems and Their Design

E.J.CRANE and C.L.BERNIER

Readiness of communication and the effort for its betterment are significantly characteristic of our times. This is one of the chief reasons for the modern good record of scientific and other progress. The tendency to label ages, as the Stone Age and the Iron Age, cannot accurately be followed contemporaneously, but surely sharing information and getting places readily have done more than anything else to make the difference between our time and earlier periods.

A proper perspective for the interrelations of the small area of communication dealt with in this paper can be given by a rapid outline of one aspect of the large field of communication in very general terms at the start.

MISCOMMUNICATION

Communication as a social phenomenon requires the cooperation of others. This well-known phenomenon is subject to abuse. Miscommunication, the negative counterpart of communication, is sometimes considered to be an offensive subject except, perhaps, in an analysis like the following. Its existence has brought into being such things as detectives, hermeneutics, polygraphs, mental institutions, and prisons. Miscommunication is probably the most common crime. Terms applied to some forms of miscommunication are: error, mistake, crime, sin, immoral, unethical, mental aberation, antisocial, psychotic, pathological, and the like, depending on the field of knowledge from which the terms are derived. Since miscommunication tends to disorganize, there is an obligation on those who wish to preserve organization to communicate properly by avoiding miscommunication. Although the subject

E.J.CRANE Director and Editor, The Chemical Abstracts Service, The Ohio State University, Columbus, Ohio.

C.L.BERNIER Executive Associate Editor, The Chemical Abstracts Service, The Ohio State University, Columbus, Ohio.

of miscommunication is much too broad for inclusion here, a sketchy outline of its scope will provide some useful thoughts related to the main part of this paper.

Intentional miscommunication is often used by self-righteous individuals or groups to control the actions and thoughts of others. Unintentional miscommunication may be caused by error, mistake, delusion, hallucination, illusion, the subconscious, etc.

Miscommunication can also be divided, for analysis, into four classes: (1) inverted communication, e.g., lying, (2) communication lack, e.g., loss of information among other information, (3) overcommunication, e.g., verbosity, and (4) concealed communication, e.g., code.

Inverted communication is often associated with words such as error, mistake, mendacity, lying, irony, untruth, fraud, bluff, misrepresentation, superstition, propaganda, slant, sophistry, rationalization, pseudotact, delusion, paranoid thinking, hypocrisy, treachery. Although some forms of inverted communication are often considered to be the most reprehensible of all forms of miscommunication, they are probably no more antisocial than species of the other classes of miscommunication. Kidding, some forms of humor, and magic for entertainment are types of inverted communication usually considered to be socially acceptable.

Communication lack (the withholding of important information) often comes about by coercion of a political, economic, or military nature, from ignorance (often of obligations), cowardice, accidie, desire for personal gain, stalling, indecision, etc. It can be caused by loss of information among other information, noise, error, etc. Loss of pertinent information among (temporarily) irrelevant or unnecessary information is one of the principal problems considered in this paper. It is often caused by sheer quantity of information and by the failure to provide adequate information selectors even though some reasonably satisfactory selectors have been devised, e.g., indexes and classifications. The general considerations in the design of documentation systems presented below will take up the requirements for adequate selectors.

Overcommunication, often used to conceal information, is associated with terms such as verbiage, verbosity, twaddle, gobble-de-gook, razzle-dazzle, padding, repetition, word-indexing, trivia, filibuster, plagiarism, nonsense, persiflage, irrelevancy, redundancy, impertinence, loquacity, sarcasm, invective, profanity, nagging, needling, chivy, henpecking, and brain-washing.

The fourth class of miscommunication, concealed communication, is associated with such terms as limited communication, jargon, cant, argot, provincialism, code, cipher, obscurantism, archaism, jamming, noise, indirection, and suggestion.

This paper deals with only a part of the field of communication and the small part of the field of miscommunication that is unintentional and possibly unavoidable.

The attitudes taken by individuals in a group toward the phenomena of communication and miscommunication have a marked effect on how the individuals are viewed and treated by the other members of the group. The scientist, who has accepted communication as a way of life and who has thereby excluded miscommunication, is often regarded as temptingly naive by those who do not take so “limited” a view of social intercourse. Scientists, who may be regarded to a certain extent as modern-day counterparts of the ancient prophets and oracles, tend, in turn, to view miscommunication as one of the principal sources of man-made troubles. The miscommunications of the business sharper, the politically unscrupulous, the tyrant, the unrestricted paranoid, the Pharisee, and the like are often viewed with contempt as a prolific source of woes and crimes. On the other hand, the “practical” people, who take a tolerant view of miscommunication, tend to regard the “eggheads” (scientists) as “ivory tower,” “long-hair,” impractical individuals who cannot “adjust” to life.

A society or organization that condones and accepts miscommunication as “normal” would seem to be placing great strain on its chances for survival. In any event, types of miscommunication, for example, about the properties of botulinus toxin, nitroglycerin, plutonium, and gravity are not conducive to the survival of those accepting the miscommunication as communication.

The borderline between miscommunication and communication is not sharp. Tact is on the borderline, and the tactlessness of scientists can often be regarded as the result of desire to avoid miscommunication as well as to preclude some confusion between tact and miscommunication. Whether miscommunication is ever justified is certainly a moot point. In dealings of psychiatrist with psychotic, slave with slave-master, police with criminals, physician with moribund, armed forces with enemy, game player with game player, and oppressed with tyrant, miscommunication comes closest, perhaps, to being justified. Whether fiction and abstract art are forms of miscommunication is an interesting point.

Just why intentional, serious miscommunication should be so socially disorganizing is not difficult to imagine. Besides the actual harm resulting from accepting miscommunication and acting on it in good faith, there is the insult felt from discovered miscommunication, which places one in the position of the psychotic, the criminal, the moribund, the enemy, the game-player, and the tyrant. There is the insecurity of knowing that this deceit may presage more deceit and conceal submerged deceit. Harm and justified feelings like these are ample reasons for the almost universal disapproval of serious miscommunication. Humorous miscommunication, e.g., kidding, joking, and fantasy, on the contrary, may lead to social solidarity by showing that the jokester recognizes the phenomenon of miscommunication and treats it with the proper contempt by making light of it, thus paving the way for mutual trust. The practice of suddenly labeling discovered serious miscommunication as kidding does, of course, not aid mutual trust.

COMMUNICATION

The field of communication (exclusive of miscommunication) is also very broad—too broad for this paper. Responsible communication of the truth, the whole truth, and nothing but the truth so far as known and possible can be divided into four principal classes: (1) religious (taken to mean that which binds the experiences of life into a meaningful whole), (2) scientific, (3) technical, and (4) inspirational (the arts). The scope of this paper, concerned in general with one kind of indefinitely delayed communication (documentary communication), is limited to communication of scientific and technical information by documents.

Communication competes for time. Time taken to watch television is not available for reading; time used to read one scientific article is lost for another. There is not enough time in a year in which to read one-twentieth of the chemical papers published in that year, let alone understand and use them. Selectivity of communication to be received is not only an opportunity, it is a necessity and a fact. As a social phenomenon, communication determines, and is determined by, the society in which it exists. In an age of communication and supermarkets it seems difficult to stay well rounded and not overstuffy.

THE STRUCTURE OF SCIENTIFIC DOCUMENTATION SYSTEMS

The documentary systems considered here can be defined as devices for communicating with unknown persons in the indefinite future. The recipients of the information may be unborn when a document is prepared. The specific information that the unknown recipients will select is also unknown. In current idiom, documentary systems are characterized by lack of feedback. Control of input by the recipients is either lacking or seriously inadequate in nearly all cases. Documentary systems, including those for captive audiences, e.g., reports and directives in an organization which do have effective feedback, are not the primary concern of this paper. The input to the systems considered here usually represents an act of faith, a casting of bread upon the water. The contributor of information to the system can have little or no tangible assurance that the results of his efforts will be found and read, let alone used. And yet communication is commonly the principal validifying outcome of his efforts in research, thinking, experiments, reasoning, etc. It is almost unthinkable that the results of the work be left uncommunicated, or that the communication be lost, at least within a limited circle, as a group of scientists, an industrial group, or a group of defense workers. From this rationally compulsory nature of communication it would seem that the act of faith must have commenced with the work that led to the communication. The work can be classed as an act of faith because the researcher, thinker, experimenter, and (or) inventor had no guarantee that his efforts would be successful.

Documentation systems can be considered to be made up of three components: (1) the contributors, (2) the storage element (including selectors), and (3) the users. It is only by interaction among the components that the systems become useful. Because of the delayed nature of the communication involved, there is a minimum of interaction between contributors and users. The interaction is largely confined to that between the storage element and the contributors or the users.

The media of interaction are multiple. Symbols, terminology, nomenclature, diction, syntax, grammar, and language are all part of the media. They are an essential part. This is true whether the storage element is manipulative (e.g., mechanized) or not. There are no other effective media of interaction known at this time. The input to the storage element consists of these media of interaction and so does the ultimate output regardless of the nature of the storage element. It is not possible to design an effective documentation system in which the primary input and the ultimate output are gibberish. Intelligible information must be put in and that kind must be recovered. The necessity of translating intelligible information into difficultly accessible machine language or symbols does not alter this fact. The number of different symbols (or symbol groups, e.g., words) that can be associated with a large collection of scientific documents is enormous and seemingly endless. This great number is a source of serious difficulty in the use of documentation systems. The problems involved in bringing vocabularies of searcher and storage-element selector into coincidence and correlation will be discussed later in this paper.

The storage elements of documentation systems, including the selectors, have been developed through the years. Libraries, publication, indexing, cataloging, classification, etc., have evolved steadily. Publication of technical articles in journals, after careful review by experts in the fields of the articles, is a well-tested technique for documentary communication. Careful editorial supervision and reviewing protect readers against several types of miscom munication. Secondary publication, of abstracts for example, is a carefully developed technique for bringing to the attention of the specialist references classified in such a way as to give him comprehensive and rapid coverage of new information in his field, no matter what the original form of communication (language, presentation, etc.) may have been and no matter how limited the accessibility may have been. Indexes, catalogs, and classifications have been designed to save the literature searcher from uneconomical, consecutive examination of documents and of abstracts to documents. They aid in gaining rapid access to the points of greatest probability for finding references to information and help to make consecutive searching unnecessary.

The technically trained individual and the scientist are aided in keeping up with most new developments in their fields of specialization by subscribing to journals devoted to these fields. The spectrum of articles in these specialized journals is usually considerably broader than the field of specialization of the reader. This is beneficial in that it prevents overspecialization and assists cross-fertilization of knowledge in a limited way. Articles are selected from these specialized journals by the tables of contents, by the author indexes, or by scanning the pages. These processes also broaden the user.

The abstract journal brings to the attention of the reader articles in journals to which he does not subscribe and a much broader range of technical and scientific papers than are found in the specialized journals which he reads. The classification of abstracts in the abstract journal helps the user to avoid abstracts that are completely irrelevant to his fields of main interest. It is probably well that the classification of abstracts be not too fine in order to provide a broader spectrum for selection and to avoid miscommunication caused by bad classification (or misunderstood classification). Even coarse classifications become obsolete, but the rate of obsolescence is much lower than for a fine classification, and adjustments are much simpler for the former. Cross references among the different classes of abstracts in the journal make unnecessary the publication of more than one abstract.

An adequate abstract journal provides an accessible, complete, permanent record in one language for its field and a key thereto. The abstract also functions as a “newspaper” to enable the user to keep up with new developments with a minimum of searching. The only partial alternatives to the abstract journal are: the compendium, the comprehensive index or classification, and the centralized information service. The first two have proved to be slow of preparation, the classification is beset by obsolescence, and the last is not universally attractive from an economic standpoint or from the standpoint of time of access or flexibility in adjustment of searching to correspond to actuality. Probably none of these alternatives can replace the abstract journal. Lists of titles of articles can be prepared rapidly, but they are inadequately useful in selecting articles of interest, and they provide little or no directly usable information. The success of abstract journals in the field of chemistry for over a century attests their importance and usefulness, as does the birth of new abstract journals patterned closely after the old. From the facts available today it seems highly unlikely that the successor or competitor of the abstract journal is in sight or even just around the corner.

Subject indexes provide the most commonly used and most effective key to information known today. The user of such indexes approaches them with a single term in mind, or with an array of related terms. The index has, classified under this term, references to any documents related to it. Thus, the index gives the user what is usually a small collection of references from which to choose. Selection from the small collection is facilitated by modifying phrases (technically called "modifications") and sometimes by classification other than alphabetical. The modifying phrases enable the reader to reject most entries that are completely irrelevant to his search. Examination of the abstracts or documents led to by the references gives the searcher the information most closely related to his search.

In author indexes usually only one term is needed to initiate and carry through the search. That term is the name of the author. In molecular formula indexes only two "terms" are needed. The first "term" is the molecular formula and the second (these are used consecutively) is the name of the chemical compound. In the numerical patent index only one "term" is needed—the patent number plus country of issue. For the subject index, in contrast, more than one term is frequently needed in making a "complete" search. It is often necessary for the searcher to make an array of related terms (1). This array will usually include general terms, specific terms, synonyms, and otherwise related terms. For example, in a search for some specific property of sodium chloride (table salt), the array of related terms needed for searching an alphabetical subject index would include: sodium halides, sodium salts, sodium, sodium compounds, alkali metal chlorides, alkali metal halides, alkali metal salts, alkali metals, alkali metal compounds, chlorides, halides, salts, and chemical compounds (2). Index headings for each of the terms in this array would be consulted. Most of the desired information would commonly be found under "sodium chloride." If the index-heading terms have been chosen during the indexing on the basis of the maximum specificity, "sodium chloride," e.g., would be selected in preference to "sodium halides," and if but one index entry related to the subject has been chosen per document (abstract, etc.), then duplication of references found in searches by use of an array of related heading terms will be at a minimum. If the array of related terms is small, then the alphabetical subject index provides what is probably the most rapid access to references leading to information. However, it may happen that the search is related to a broad, generic question, such as, "What are the biological properties of chlorinated hydrocarbons?" Since there are hundreds of biological properties and potential billions of chlorinated hydrocarbons, the construction of an array of search terms is impossible because of lack of time. Thus, for broad, generic questions, the alphabetical subject index is not so useful as a classification that brings together, for example, names of all chlorinated hydrocarbons associated with biological properties in documents. Since the potential number of classes and combinations of classes about which information *may* be sought is astronomical, it is impossible to design a hierarchical classification that includes *all* classes and *all* combinations of classes. The designer of a hierarchical classification is compelled to choose which of the classes and combinations of classes he believes will be most important.

The ratio of numbers of questions involving specific searches to those involving broad, generic searches is unknown, as is also the effect of comprehensive searching means upon the ratio. It seems likely that, if infinite flexibility of search were available, then broad, generic searches would become much more frequent than they seem to be at present.

Fairly recently, manipulative correlative documentation storage elements have been discussed extensively (3). All these storage elements have, in effect, selectors based upon nonhierarchical classification (4). Correlation of terms makes increased selectivity possible. The user of these systems, which may be called "correlative indexes" (5), usually selects two or more terms related to his search. The use of two or more terms simultaneously gives a very powerful selecting effect. The more terms used simultaneously the greater is the selectivity. In fact the selectivity is so great that there is serious danger of loss of part or all of the information if four or more terms are used simultaneously (6). Most correlative indexes discussed up to the present have been manipulative, i.e., terms selected by the searcher from the vocabulary usually are subject to the manipulation of correlation in order to recover documents or references to documents from the storage element. Many of the correlative systems have been "mechanized," e.g., by the use of punched cards, microfilm, and magnetic tape.

PREDICTION IN DESIGN OF DOCUMENTATION SYSTEMS

Since delayed use of documents and the anonymity of the user are usually expected, prediction is necessary in the design of documentation systems. Without prediction (or extrapolation), design is often impossible because of lack of feedback and of specific knowledge of the needs of users—some unborn. The requirement of prediction may make the design of documentation storage elements seem impossible; yet there is apparently no other way in which to design them. However, since actual systems function fairly effectively, and have done so for at least a century (7), the barrier of prediction is proved to be passable. As with any prediction by extrapolation, it seems likely that short-range predictions will probably have greater accuracy than long-range ones. This is especially true in the design of classification systems which are prone to the ravages of obsolescence.

Although prediction, when viewed as a whole, may seem so hazardous as to dismay the designer, it turns out to be a relatively safe procedure when viewed in parts. It seems reasonably certain, for example, that the alphabet and order of letters therein will remain fixed for many centuries to come as they have for many centuries past. These two reasonable possibilities are very good fortune for the designer of documentation systems. They mean that the designer can probably rely on alphabetical order as a guide to terms in the vocabulary of the selector and hence to documents in the system for the user in the distant future. Also, the system of numbers and decimals, and their order, seems to be very durable. Numerical order seems likely to be preserved through the centuries. Alphabetical and numerical order are of great, if not of vital, importance in the location of documents and of references to them in all known documentation systems (including those mechanized); it seems likely that they will continue to prove important in all future documentation.

Another prediction that can be made with reasonable certainty is that the nomenclature and terminology in a given field will not often be altered suddenly or radically (8). Most changes usually will be the gradual additions of new names and terms and the dropping of obsolete ones. Such changes can be accommodated easily in indexes by cross references together with the indication of synonymy and other relationships.

Another durable feature of documentary systems seems to be grammar. Abrupt changes are certainly unexpected. The language of today seems likely to be easily understandable in the future.

In the design of documentation systems for future use, it seems necessary to assume that the language, terminology, nomenclature, and symbols employed will be understood by the user, can be made understandable to the user in some simple manner, or can be generated by simple rules (9).

The input to the storage element in present-day documentation systems is already largely predetermined. The quantity, content, terminology, nomenclature, diction, syntax, grammar, symbols, etc., of documents are, in most cases, beyond the control of the designer of the storage element. He can alter the arrangement and form of the documents to facilitate storage, but he must not miscommunicate by altering meaning. He can, for example, select, abstract, extract, classify, index, translate (perhaps into "machine language"), and abbreviate as well as rearrange in many cases. The information lost in these operations must always be "trivial" or "irrelevant." None of these operations is as complicated as it might seem on first thought. Rules for performing most, if not all, of the operations have been developed successfully and recorded. The builder of the storage element and selector cannot usually *add, subtract,* or *alter* (in certain ways) information in the system without, at least, some miscommunication. The fact that the builder must generally accept what he is given does not prevent him from checking in reference works and with the author of a document about errors discovered. To this extent feedback is practical and important in documentary systems.

Since the builder of the storage element and document selectors cannot include *all* documents in the collection and cannot use *all* words and other symbols in abstracting and extracting, he must select. Also in indexing and classifying, selection is desirable and necessary. In translation, on the contrary, selection of parts of documents is usually avoided. In making selections of documents, index entries, members of classes of a classification, etc., the builder is knowingly or unwittingly controlling the effectiveness of the system for the future user. The predictions that he makes about what the future user will need are the principal factors controlling his selections. There are a number of bases for selection: Documents can be selected as being related to a given field of information. They can be chosen by form, e.g., patent specification, technical paper, book, separate, classified information, unclassified information. They can be picked for inclusion in the collection by their novelty. They can be selected to cover a certain time span, e.g., one year. If funds permit, an attitude of generosity can prevail in the selections.

The indexer can choose index entries by emphasis and novelty (10). He can index subjects or select only words and thus produce a concordance or partial concordance. He can choose special classes of material for indexing, such as molecular formulas for chemistry, names of authors, patent numbers, topical subjects, geographical locations, time spans, etc. He can index terms of the maximum specificity as far as possible. The classifier can choose many bases for classification. The builder of a correlative system for selecting documents can choose terms for the vocabulary. These terms can be general, specific, or both. Terms associated with certain phases or classes of information can be chosen on most, if not all, of the bases indicated for indexing. If the indexer also includes trivia and irrelevancies, the useful index entries may become so diluted as to be subject to loss. This is true for correlative indexing as well as for alphabetical subject indexing, although probably to a lesser degree. Whether terms selected from the documents in the collection are to be used in an index in book form or in a manipulative index, the above concepts are still applicable. The selection process and rules remain largely unassociated with the mechanism of the selector. It is important to note that the success of the documentation system in providing relevant documents to the future user is determined, up to this point, by the selection of terms by the indexer and not by the nature of the selector. If the term selection rules are inadequate or inadequately observed, then miscommunication will result.

The abstractor or indexer (for any type of storage element) has no choice but to use symbols (usually words) found in the documents, the symbols which are synonyms of these symbols, or terms semantically related to these symbols. He cannot choose terms completely unrelated to the documents processed without complete miscommunication. The important point here is that the abstractor and indexer are greatly (and properly) limited by the document abstracted and indexed. The meaning of the document was fixed by the author; the abstractor, indexer, documentalist, or other person cannot change the meaning without miscommunication. The number of terms, their synonyms, and semantically related terms that can correctly be associated with any document (except dictionaries, etc.) is very small if compared with the total number of terms available in a language. About the only liberty that the chemical indexer can have is that of changing the names of chemical compounds to synonyms which conform to an internally consistent systematic nomenclature harmonious with previous indexing.

As pointed out above, classified, thoroughly indexed abstracts in one language provide one of the most effective solutions to the problems of the many languages in which technical information is written and of the dispersity and volume of the publications. This solution to the problem gives the future user the present-day airliners of documentation. Whether jet planes of documentation are just over the horizon is a moot point. Certainly some individuals have had the impression that all information could and should be recovered at the press of a button. They have largely neglected to insist on wide use of techniques already proven fairly satisfactory—at least until proven new methods were available. While waiting, hoping, and working on push-button service will it not be wise for these individuals to discover, more fully understand, and more assiduously utilize the methods developed up to now? These methods serve the purposes of the literature searcher much more effectively than some seem to realize.

The selection of abstracts for inclusion in an abstract journal is associated with the same problems as is the selection of documents for inclusion in a collection. In addition there are the problems of what to select for inclusion in the abstract. Something must be omitted. Rules for effective abstracting are derived, fundamentally, from predictions as to needs of future users. Adequate rules for abstracting and careful and consistent application of these rules assist communication with the unknown user. It is interesting to note that rules for effective abstracting have been in existence for at least a century. When the rules fail, or are ignored, then miscommunication may occur.

The abstracts can omit historical detail since the assumption can usually be made that earlier abstracts have already included this information. Abstracts can exclude also much specific numerical information, tabular material, graphs, complete mathematical proofs, and usual experimental details—and include, perhaps, only the highlights and maxima and minima. In other words, abstracting can probably always be successful if carefully done. The evidence of at least the last century backs up this statement. It is good to know that the designer of future documentation systems has this proven resource to aid in his predictions and design.

Classification of published abstracts has assisted the user in their selection before subject and other indexes were available. The existence of cross references among the classes has further aided in this. If duplication of abstracts is prohibited in a given abstract journal, then cross references among the classes is about the only way of avoiding scattering of like information before the indexes are published. Difficulties with classifications are sometimes experienced in the placing of abstracts in an abstract journal. Scattering of like abstracts may occur. Abstracts are often not distinct units as to subject matter, and they sometimes fit into more than one classification about equally well. However, classification is definitely an aid. Without classification, selection would in most cases be much less rapid.

It has been suggested that abstracts be written or translated into a very rigid form suitable for searching by machine. This type of abstract has been called “telegraphic” (11, 12). Some experimentation has been done in this field. The rigid form is based on prediction of future needs and imposes on the documentation system additional rules which have been derived from the past. Certain fields of knowledge, e.g., preparative organic chemistry and metallurgy, may be better adapted to this technique than other fields, such as theoretical chemistry, because of its extensiveness and unpredictable variations.

It is obvious that a document omitted from a collection is not available to the future user and that an index entry omitted will make a document unavailable from one indexing viewpoint. Miscommunication may result from these omissions. If too many irrelevant documents are included and too many ir relevant index entries are chosen, then another type of miscommunication may result, namely overcommunication.

All the selections discussed above are based upon intelligible symbols, usually words, according to the meaning given by the context. Words used as the basis for selecting, indexing, and classifying a given document may not appear in the document. Judgment is required. The basis for sound judgment is a sound background in the field of knowledge associated with the collection together with training in selection, indexing, and classification. The most costly part of the construction of a documentation storage element is usually the hiring of judgment needed for making these selections consistently, accurately, and comprehensively. The less costly part is usually the publication of these selections. All these selections presuppose that the future user will find them understandable, significant, and useful. The selections will probably be understandable to the user because of the continuity of language and the semantic durability of vocabulary terms (9) mentioned above. They will probably be significant to him if the indexing or classification is based upon subjects (not words), novelty (excluding old, well-established information which has been adequately documented and indexed before), emphasis (excluding trivia), and a common background of knowledge.

Selection by the user will be effected by symbols (usually words) that he knows, that can be brought to his attention, or that he can generate by simple rules, together with an adequate dictionary or thesaurus for explaining the meanings of any symbols that he does not understand. In order to bring all significant, relevant information to the user, he must have convenient access to the documents in the collection and to all the words in the selector vocabulary that are pertinent to the question asked or to the general area in which it is hoped that information will be discovered.

It is necessary that the user be aware of the existence of the collection of documents. It seems most likely that this awareness will come about initially through another individual, e.g., relative, teacher, librarian, colleague, supervisor, as the first source of information about the collection or about publications leading to knowledge about the collection. As a secondary source, publications about collections of documents (13) are exceedingly valuable.

In order to operate the selector devices (indexes, classifications, mechanisms, etc.) the user must be able to select words pertinent to his problem from the selector vocabulary that was used to index the collection of documents (14). If this selector vocabulary is small—say with fewer than one thousand terms— then complete reading of it is probably the most effective way of bringing the pertinent vocabularies of searcher and selector into coincidence. If the vocabu lary is much larger than this, then some form of classification or thesaurus is probably necessary to effect this coincidence rapidly enough. Hierarchical classifications suffer from obsolescence—as mentioned above—because they are built on existing knowledge and so may not be able to handle what is new. The designers of classifications, lacking omniscience, cannot anticipate completely new categories and their interrelationships. If the documentary system is to have a life of, say several centuries, it now seems most probable that, with the extraordinarily rapid expansion in knowledge (especially in the fields of the sciences), hierarchical classification would prove totally inadequate in organizing the vocabulary terms for effective discovery in the distant future. The history of hierarchical classification seems to bear out this statement. Continuous revision, promptly accepted and published, seems to be about the only method of avoiding or of minimizing obsolescence. Revision has usually proved to be costly and slow.

A more promising approach to the bringing of large vocabularies into coincidence seems to be the technical thesaurus (2), which enables the user to go from the words that he knows to all of those that he needs to know for a complete search. Such a thesaurus, which somewhat resembles a hierarchical classification, but is multidimensional and comprehensive, is also faced with obsolescence. The task of maintaining it current seems to be much less formidable than that facing the builder of the hierarchical classification. New words are incorporated as they appear. Their relationships with words already in the thesaurus will usually be obvious or readily discovered and easily incorporated, and will usually not disrupt relationships already established. The thesaurus defines, in effect, the terms included and thus helps to make a dictionary less necessary. It now seems possible to use the thesaurus to derive the very general terms for use in building the small vocabulary of a document-selecting system. These general terms should probably be selected before correlative indexing is started. If general terms are added to the vocabulary during the construction of a correlative (e.g., mechanized) index, then the indexer must review his earlier indexing to check it for the use of the additional terms in indexing. This kind of review does not seem so efficient as the building of the thesaurus first, maintaining it current, and using the generic terms discovered for correlative indexing. By way of contrast, the alphabetical subject index, built with heading terms of the maximum specificity, does not have this kind of problem. New words are incorporated normally as they appear. Cross references may be used to relate the new words to others already incorporated. As pointed out above, the principal difficulty in use of the alphabetical subject index lies in making generic searches in which the array of specific headings under which to search is too large or indefinite for effective selection. Correlative indexes (manipulative or nonmanipulative) seem to offer hope of greater effectiveness in generic searches (and possibly in specific searches also), especially if the vocabulary can be kept small and (of necessity) generic.

OUTPUT OF DOCUMENTATION SYSTEMS

Not all forms of output of documentation systems and their contents are equally desirable, available, and rapid. Certainly no documentation service can surpass in speed that in which the precise document required is selected and ready before the question is asked. It is interesting that services coming close to this ideal have been developed. The Library of Congress system (15) for bringing current information to members of Congress and SVP (16) in Paris are two examples.

Next most convenient are services which provide a small collection of documents from which the searcher can choose those of greatest importance to him. Libraries provide this service for books as a matter of course. The order of books on shelves provides this service almost automatically.

Less convenient is a selected bibliography; still less is a general bibliography. A selected list of references without titles and a general list follow in convenience.

Indexes and classifications by means of which a bibliography or list of references is generated come next. A documentation center at a distance which might be approached and reapproached by letter, telegraph, phone, or teletype as ramifications of the question are explored is probably the least convenient of the documentation services. Such a center seems to offer less opportunity for the searcher to grow in his search (17).

The ultimate (and often penultimate) output of all documentary systems must consist of intelligible symbols. The output will usually consist of words, etc., on paper, photographic films, etc. Documents, references to documents, bibliographies, etc., may be obtained as the result of a search. The references to documents and bibliographies will usually be on paper or film as an aid to memory. The output will usually be the same whether the storage element and selector are manipulative or not. The usual response to a question asked of a mechanized system, for example, will be in the form of marks on paper or images on photographic film, if not the actual documents or reproductions of them.

The builder of the selector (index, classification, etc.) must predict, as discussed above, what types of information may be needed. Although the designer of a manipulative correlative index does not need to predict what combination or permutation of vocabulary terms the future user will select, he must predict the type of terms most probably useful as well as the best rules for selection of the actual terms used in indexing each document. It is these predictions that may be a major controlling factor in the future use of the system rather than the structure of the storage element and selector. If the designer omits terms from the vocabulary or omits associating vocabulary terms with documents, then miscommunication may occur, whether the omissions were intentional or inadvertent.

It has been thought that publication of all effective combinations and permutations of terms from the vocabulary of a manipulative documentation storage element selector was impossible because of the size and cost of the resulting publication. However, by the exclusion of potentially useful combinations of terms and the inclusion of only actually used combinations, by the use of alphabetical order to reduce the number permutations, and by the use of a few permutations and syndetic devices to control the number of combinations, it is now believed possible to produce nonmanipulative correlative indexes in book form or in the form of an unpunched-card file. One technique for doing this has been described (5). It now seems to be possible to give the future user the results of all significant searches in the documentation system without the necessity of doing the manipulative searching. That is, the future user can be given the marks on paper without the necessity of owning or renting a mechanized system in order to obtain the marks. In the design of such nonmanipulative correlative systems it is not necessary to predict which combinations and permutations the future user will select because all valid ones are preselected and published.

It might turn out, if the above principles were not understood or believed, that a mechanized documentation system could be set up with the following results: The capacity of the machine would probably not be the same as the demand. Probably the machine or machines would be able to supply answers to more questions than were asked. The idle time of the machine(s) would be put to good use by providing answers for preselected *types* of questions. That is, the answers (or often references to the answers) would be ready as marks on paper before the questions were asked. Thus, nonmanipulative correlative indexes (with the answers to preselected types of questions) might come about from the possibly mistaken belief that mechanized documentation was necessary or inevitable.

From the above discussion it seems that the structure of the storage element is not of greatest importance in the performance of documentation systems. It is the interaction between the contributor, and especially the user, and the storage element that seems to be of most importance. The predictions in the design of storage elements and selectors for manipulative (e.g., mechanized) systems seem, in this analysis, to turn out to be the same as those for non manipulative systems (e.g., indexes and classifications). The success of these predictions determines, in both cases, the effectiveness of the systems for the future user. If predictions about what the future user needs are inaccurate, then it seems likely that both manipulative and nonmanipulative document selectors will prove, at least to a certain extent, inadequate.

So far as economics and efficiency are concerned, there are probably great differences between manipulative and nonmanipulative systems. The selection and acquiring of documents for both systems would be equally costly. The cost of indexing is probably close to the same for both systems if equal precision is achieved. If modifying phrases are not used in the manipulative system, the indexing might be less costly. The cost of the storage element of the selector for the manipulative system is likely to be much greater.

THE MAJOR DIFFICULTIES

Twelve major difficulties with manipulative documentation systems have been discovered (5). Most of these difficulties are associated with some form of unintentional miscommunication.

1. The imperceptible loss of relevant information caused by correlation of too many terms (and (or) the wrong terms) simultaneously. The searcher may not know and may not be able to discover which of the terms that he has selected as pertinent to his problem must be excluded from a given combination of terms used in searching. The difference between too few terms and too many in searching seems to be between one and four. The difference is unpredictably variable. Additional terms which are redundant, representative of a series of questions asked simultaneously, and omitted as “old information” by the indexer must be excluded in searching; otherwise selection of significant information will be blocked. The searcher interested in discovery rather than recall of information may not have enough knowledge of the new field to enable proper exclusion of terms or proper combinations. It now seems probable for questions involving discovery rather than recall that the searcher should not expect to receive a precise answer to his question. Instead, he should always expect to receive related or analogous information—some of it very closely related and closely analogous. For questions of recall, the searcher should, apparently, always expect to receive the pertinent information or document(s). The reasoning behind these conclusions is elaborated elsewhere (6). From these results, it may be useful to picture the use of a collection of documents in this fanciful way. The documents are stored on one long shelf. The searcher sits near one end of the shelf. In response to a question by the searcher, *all* the documents on the shelf are rearranged (perhaps by a subject expert, machine,

etc.) so that the document or documents most relevant to the question are at the end of the shelf nearest the searcher. The farther the document from the searcher the less its relevance to the question asked. The searcher starts by examining or reading the document closest to him and continues by consulting documents along the shelf farther and farther away from him until he reaches a point at which he feels that he is obtaining too little relevant information to continue the search. In response to a second question, *all* the documents on the shelf are again rearranged to bring them into an order of increasing relevance to the second question, with the most pertinent documents again nearest the searcher. This fanciful picture of the use of a collection of documents or a library is not intended to be a preliminary sketch for an actual library. It is presented solely to illustrate several postulates which seem to be of importance in the use of any collection of documents. These postulates, which seem backed by the experience of librarians, documentalists, designers of machine documentation systems, and the like, are:

a. Documents in a collection show different degrees of relevance to a given question.

b. Probably all documents in a collection are related, albeit some very remotely, to a given question.

c. The order of relevance of documents in the collection to different questions will nearly always be different.

d. For questions of discovery it will be exceedingly rare that only one document will be worth the time of the searcher for examination. It will usually be that more than one document will be of considerable relevance to the question.

e. For questions of recall, it may often turn out that only one document is all that is required.

If these postulates be true, then it means that the searcher who has asked a question of discovery should be given at least a small collection of relevant or analogous documents. Only very rarely will it be possible to pinpoint the information which he requires. This is so, because it is highly probable that precisely the information that he requires has never been published.

2. Blank sorts (6) resulting from searching for nonexistent classes. The number of combinations of terms taken from the vocabulary of a fair size, four or more at a time, is enormously greater than is the number of documents in even a very large system. Thus, many of the logical combinations of terms will be associated with no documents. If these combinations are used in searching, then blank sorts will result. A genuine blank sort is a desirable outcome of many searches. However, distinguishing between genuine blank sorts and the imperceptible loss of relevant information described above may be difficult or

impossible for the searcher with too little background in a new field that he wishes to enter. Time for a search terminated by a blank sort is lost.

3. The unavoidable selection of unwanted, irrelevant information. Such irrelevant information has been called “false drops” and “noise.” If the incidence of use of terms in indexing is high and too few terms are used in searching, then the number of irrelevant documents may become a flood (overcommunication) that makes selection of the pertinent documents time-consuming. For example, if the incidence of use of a term is ten percent in the indexing of a million documents, then the use of this term alone in searching will result in the selection of one hundred thousand documents. The use of two such terms might select, on the average, about ten thousand documents.

4. Confusion of meaning because relations among the vocabulary terms are difficult, if not impossible to show completely, i.e., many of the systems lack morphemes. Venetian blind *vs.* blind Venetian, man bites dog *vs.* dog bites man, and reactions in benzene *vs.* reactions of benzene are a few simple examples of this confusion. Such confusion results in the selection of irrelevant information.

5. Deficiency in effective and immediate suggestion of closely related and substitute information. According to statistics on the blank sort and for questions of discovery rather than recall, one should not expect to obtain, as mentioned above, precisely the information he seeks from any documentation system, manipulative or nonmanipulative. One should expect, instead, to obtain a small collection of documents (or references to them) closely related to the question asked. With manipulative systems it may often be difficult or impossible to obtain such a small collection. The user may feel that he is flying blind in the system and lacks adequate control of search.

6. Necessity for manipulating the system before relevant documents, etc., can be located. That is, the results of all significant correlations are not immediately available without manipulation. The time required to program and to manipulate the selector device, e.g., operate the machine, is probably usually greater than the time required to locate preselected combinations in a nonmanipulative correlative index, e.g., in book form. The time of search which results in a blank sort in manipulative systems is lost. In questions involving discovery rather than recall, the loss of time from blank sorts can prove incapacitating under certain conditions.

7. The relative bulkiness of the recording media and associated apparatus when compared with indexes in book form. The bulkiness of indexes, sometimes criticized, is usually much less than that of punched cards and the like.

8. The costliness of manipulative systems described and dreamed about to

date makes their use seem even less attractive in view of all of these other problems.

9. Delays caused by the necessity of manipulating the system or of communicating with a "documentation center" in order to get the answer to a question. For many searches nonmanipulative systems would be much more rapid.

10. The economic restrictions imposed by the more costly systems which reduce the total amount of information communicable. If the price of a storage element or of access to it is increased, then fewer users can be helped by it. Communication lack results.

11. The bringing of pertinent vocabularies of searcher and selector into coincidence or correlation. This is not exclusively a problem associated with manipulative indexes, although it may be more acute with them because the failure to effect correlation in this case may give irrelevant data or blank sorts, and may give insufficient information to the user to enable him to correct his searching technique.

12. The facilitation of generic searches. This is another problem that is also associated with nonmanipulative indexes. A small vocabulary of generic terms and the technical thesaurus may be helpful here also.

At present, there seem to be about five distinctly different ways of effecting coincidence of vocabularies:

a. The use of the documentation system by the builder. A "one-owner" system should experience little difficulty from this source since the builder uses his vocabulary for recovery of information from the system as well as for indexing it. His memory largely solves the problem of coincidence. This way should also apply fairly well to a documentation center built and searched by the same staff.

b. The use of rhetorical tropes of a generic nature. This technique, which has been pioneered by Mooers (18) through the use of his generic descriptors, is described from the viewpoint of rhetorical tropes in a recent paper (14).

c. The use of a comprehensive thesaurus which enables the searcher to go from the words that he knows to all words necessary for the search. A thesaurus of this nature, which has been described (2), should have some very interesting auxiliary properties, such as functioning as a current classification system for a given field—a system easily avoiding the ravages of obsolescence.

d. The generation of new names and terms by the use of rules. Notations (ciphers) (19) and systematic organic nomenclature (20) are two examples of the use of this technique for bringing vocabularies into coincidence. Rules, instead of words or symbols, are communicated by the builder of the documentation system to the user. The vocabulary of the searcher is brought into existence by rule, as required. It is interesting to speculate on what fields other than those of organic chemistry and radio tubes can be handled by similar systems. It may turn out to be that every class of things with structure and measurable properties can be so handled; e.g., maps, radio circuits, blueprints, insects, stars, pictures, flowsheets, machines, graphs, animals, and plants.

e. The use of auxiliary aids (other than a comprehensive thesaurus). Such aids include conversation with colleagues, dictionaries, cross references in the documentation system, courses of study, encyclopedias, compendias, reviews, papers, and references.

SUMMARY

Intentional miscommunication (except in experimental psychology, etc.) is incompatible with the documentation of science. Unintentional miscommunication can be reduced by established procedures. Design of documentation systems involves prediction because the users are unknown and the communication is indefinitely delayed. Selection of documents, index terms, entries, and classifications involve prediction of probable uses. There are a number of good bases for prediction: It has proved possible to define a field of knowledge fairly precisely and actually to use this definition in selection of documents to be included in a collection. The existence of successful abstract journals for at least a century is proof of this. It seems reasonably certain to predict that the definition of field will remain effective for a long time, perhaps another century. It seems possible to predict that the numbers and letters now in use and their present order will, a century from now, be substantially the same if not identical. This prediction makes practical the use of lists in alphabetical and numerical order. The prediction that the structure of language, grammar, syntax, etc., will probably be the same for a long time enables one to use current language, grammar, syntax, etc., with confidence that miscommunication to the future user will probably not result. The prediction that nomenclature and terminology will usually shift only gradually and that devices such as cross references will be available to avoid loss of the older terms enables the builder of the system to use the nomenclature and terminology of the present day in the selector element of the documentation system. The prediction that the future user will be interested mainly in novel (at the time of storage), emphasized subjects (not words) seems reasonable from the experiences of about a century in the indexing of abstract journals. Predictions as to needs for specific versus generic information are more difficult to make. It seems likely now that both types are required, and that the emphasis and importance of each are the only doubtful points. If the designer of a documenta

tion system for strangers in the future believed that the strangers would not have an alphabet, system of numbers, etc., then he would see little point in doing work on the system.

In conclusion, the documentation systems for the future must be built with the information that we have today. They should be capable of absorbing *new types* of information without bursting their classifications, or rather, the classifications should be capable of indefinite expansion with a minimum of effort spent on maintaining them current. The alphabetical subject index, the numerical patent index, the author index, and the molecular formula index are four document selectors with just these properties. At the present time it seems probable that the results of all meaningful searches in mechanized systems can be prerecorded in economical indexes before the questions are asked and thus save the time and money spent in developing machines and in using them. The economics and logistics of documentation storage elements seem to be heavily in favor of indexes published in book form. Techniques for improving the speed and reducing the cost of indexing and of printing indexes are desirable. Convenient (almost automatic) growth by the searcher during his searches makes nonmanipulative indexes seem even more attractive.

The insertion of a machine as storage element and (or) selector between contributor and user does not solve the major problems of documentary systems because these problems are associated with the growth rate of the scientific literature and with the interaction between storage element (including selector) and contributor or user. These problems, which are related to growth rate and to selection and prediction of documents, index entries, and vocabulary terms, are almost entirely unaffected by the introduction of selection and storage machines. On the contrary, the introduction of machines as storage elements and (or) selectors has brought some of the serious problems discussed above. The effort invested over the last decade or so in ingenious mechanized systems has not been lost, however, for it has served to point up the powerfully selective properties of correlative indexes [described in the early nineteen hundreds (21)] and to produce actual systems which work rather well under limited conditions, e.g., for use by the builder and (or) for rather small collections of documents relating to limited subject matter. The speed, tirelessness, and accuracy of machines have been powerful stimulants to thinking along these lines.

REFERENCES

1. C.L.BERNIER and E.J.CRANE, Indexing abstracts, *Industrial and Engineering Chemistry, 40,* 730 (1948).
2. C.L.BERNIER and KARL F.HEUMANN, Correlative indexes III. Semantic relations among semantemes—The technical thesaurus, *American Documentation, 8,* 211–20 (1957).
3. R.S.CASEY and J.W.PERRY, *Punched Cards, their Applications to Science and Industry,* Reinhold Publishing Corporation, New York, 1951.
4. LEA M.BOHNERT, Two methods of organizing technical information for search, *American Documentation, 4,* 139 (1955).
5. C.L.BERNIER, Correlative indexes I. Alphabetical correlative indexes, *American Documentation, 7,* 283–8 (1956).
6. C.L.BERNIER, Correlative indexes V. The blank sort, *American Documentation, 9,* 32–41 (1958).
7. *Chemisches Zentralblatt* was started in 1856.
8. LEONARD BLOOMFIELD, Linguistic aspects of science, *International Encyclopedia of Unified Science* Vols. I and II, No. 4, p. 8 (1950), University of Chicago Press, Chicago, Illinois.
9. C.L.BERNIER, Language and indexes, *American Documentation, 7,* 222–224 (1956).
10. C.L.BERNIER and E.J.CRANE, Indexing abstracts, *Industrial and Engineering Chemistry, 40,* 727 (1948).
11. G.M.DYSON, Studies in chemical documentation III. Mechanized documentation, *Chemistry and Industry, 1954,* 440–9.
12. New tools for the resurrection of knowledge, *Chemical and Engineering News, 32,* 868–9 (1954).
13. E.J.CRANE, AUSTIN M.PATTERSON, and ELEANOR B.MARR, *A Guide to the Literature of Chemistry,* John Wiley and Sons, Inc., New York, 1957.
14. C.L.BERNIER, Correlative indexes II. Correlative trope indexes, *American Documentation, 8,* 47–50 (1957).
15. C.A.GOODRUM, The reference factory, *Library Journal, 82,* No. 2, 121–30 (1957).
16. *Realities* (The French magazine printed in English) p. 37, April 1957.
17. C.L.BERNIER and E.J.CRANE, Indexing abstracts, *Industrial and Engineering Chemistry, 40,* 730 (1948).
18. CALVIN N.MOOERS, Zatocoding applied to mechanical organization of knowledge. *American Documentation, 1,* 20–32 (1951).
19. G.M.DYSON, *A New Notation and Enumeration System for Organic Compounds, Second Edition,* Longmans, London. This reference in lieu of the final publication of the International Notation.
20. Introduction, with key and discussion of the naming of chemical compounds for indexing, *Chemical Abstracts, 39,* 5867–5975 (1945).
21. EUGENE GARFIELD, private communication.

The Possibilities of Far-Reachmg Mechanization of Novelty Search of the Patent Literature

G.J.KOELEWIJN

The Netherlands Patent Office has undertaken research into the possibilities of far-reaching mechanization of novelty search in the patent literature, starting with the subject, *carburettors*. When the outline of a paper relating to the carburettor study was drafted, it was expected that by the time the paper itself was to be submitted a preliminary test would have been performed with the available material. Unfortunately this has not been possible. Barring unforeseen circumstances we are quite confident that within 3 to 4 months, i.e., well in advance of the conference, the system will have been put to a test. In view of the special nature of the selection means that have been developed during this research, I am of the opinion that despite our failure to report the final results achieved, a paper describing our work could be important enough for the conference, and I have decided to submit a report on our progress to date.

AIM OF THE CARBURETTOR PROJECT

The most important aim of this study is to obtain a better insight into the possibilities and difficulties with regard to mechanical selection of patents for the purpose of the novelty search. The desirability of such studies for the U.S. Patent Office is admirably brought out in the Bush report (1). We were of opinion that a mere study of the literature on the subject would be quite inadequate for the purpose. It was considered necessary to get some experience of our own. In view of there being only a few members of the staff available to pursue this object, a certain choice had to be made. There were several reasons for choosing a subject in a mechanical engineering field:

1. In this field the shortage of staff was least in evidence.

G.J.KOELEWIJN Member of the Board of the Netherlands Patent Office, in charge of the Classification Department and the Library, The Hague.

2. Mechanical topics are as a rule more intelligible to those who are inexperienced in the given field than chemical or electrotechnical topics. Experience is thus more easily gained.
3. There already exists an extensive literature on mechanical selection in the chemical field, but there is very little relating to mechanical or electrotechnical topics.
4. The problem of mechanical selection of electrotechnical circuit diagrams appeared too difficult to begin with.

Carburettors were chosen for the following reasons:

1. In regard to this field we are convinced that by the usual documentary methods it is not possible to achieve a fairly extensive subclassification that is really satisfactory.
2. The aspect “shape” which is difficult to define for purposes of mechanical selection plays no important part here.
3. Here as with many other subjects that are difficult to classify the emphasis lies in the constant combination, in different relation to each other, of a limited number of detailed characteristics that are not very special when considered separately.
4. From the point of view of classification there appears to be a high degree of correspondence with electrotechnical circuit diagrams.

DOCUMENTARY SYSTEMS IN USE AT THE NETHERLANDS PATENT OFFICE

THE SYSTEMATIC COLLECTION OF PATENTS

At the Netherlands Patent Office the systematic collection of patents comprises 5 million patents (Belgian, British, French, German, Luxembourg, Netherlands, Swiss, and U.S.). Since 1946 other publications or abstracts thereof have also been incorporated in the collection. All these documents are arranged according to the headings and subheadings of the Netherlands classification system. In principle, each document is classified under the symbols of one heading or subheading, but may be allotted symbols of one or more other (secondary) headings (or subheadings). On the average, about 130 secondary headings are in use for every 100 Netherlands patents; at present about 40 secondary headings are in use for every 100 recent foreign patents, the extra copies called for being ordered specially or reproduced photographically. One great advantage of such a collection is that after determining the headings and subheadings of the system relevant to a novelty search, the examiner can at once consult the patents and other documents filed together in the appropriate filing cupboards. The classification system for such a collection in which as far as possible each document is filed under a single symbol, is, naturally, to a high degree adapted to this purpose. The same applies to the mode of classification of the patents—attention is focused primarily on the novel concept disclosed therein.

Since the very purpose of the existence of a patent is to disclose and define a single inventive and/or novel concept, an experienced worker can as a rule quite rapidly form an impression regarding the essence of the invention, for which purpose usually only one or a few points of view are of real importance. In consequence classification of such patents in the Netherlands system does not take much time. On the average, well-qualified personnel, with experience of classification work, require only 7 minutes per patent for the complete operation of classification. Considering that the Patent Office acquires each year some 150,000 new patents that are incorporated in the classified collection, the classification of these demands the equivalent of the annual working time of 10 experts (the actual number engaged on this work, as part of their duties, being far greater).

Many objections can be levelled against this system. On the other hand it has the great advantage that it requires relatively very little time of technical experts and that the patents themselves are arranged in systematic order.

In all technical fields in which with this system a novelty search continues to afford reasonable results in a reasonable time, it is apparent that the introduction of any other system cannot be justified on economic grounds. Here one has to take into account the annual total of patent applications; what may be uneconomic for our patent office may well be feasible for a much larger organization or in the event that another system is brought into being on an international scale by a cooperative effort of a number of organizations.

One of the most obvious objections to a system in which the patents are arranged in groups is that the steady growth of material (there is constant evidence that old patents can be important) makes constant subdivision of the groups necessary, so that in most cases the limit has already been reached at which the subject matter of a patent falls completely within the scope of a particular group of lowest hierarchy. Moreover, classification according to the (presumed?) inventive concept involves a limitation that has certain practical advantages, but it is nevertheless a serious limitation, for it results in much, sometimes very important, information contained in a patent being to all intents and purposes unavailable for purposes of novelty search. After all, what matters for a novelty search is not particularly the scope of the claims in the published patents, but everything that appears therein. One will of course always have to make a choice in regard to what information is to be traceable via the documentation system, since the incorporation in a traceable fashion of all kinds of information that is already well known from other sources is of no value in itself, and can easily lead to overdocumentation.

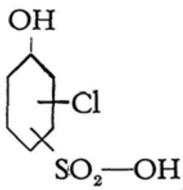
SPECIAL SYSTEMS

In any event, there has for a long time existed a need for multiple classification. In the case of a subclassification within a small self-contained group of patents (e.g., a group that has become inconveniently large) this can be achieved without sacrificing the possibility of arranging the patents themselves; the latter are analysed in the light of the subdivisions envisaged and are marked with one or more symbols accordingly. Patents similarly annotated are then grouped together and these various groups are then distinguishably arranged in a definite order in accordance with a chosen hierarchy for the symbols and their combinations. This little known possibility has turned out to be most practical in certain cases.

If the multiple classification extends to a rather large number of aspects, separate auxiliary means are called for. Since about 1930 we have made use of what we call combined lists in the fields of automatic telephone systems and of electrical control mechanisms. These lists include a column for patent numbers, a column for the classification symbol under which a patent falls in the normal system, and a large number of columns in each of which the presence of one or more aspects can be indicated by appropriate symbols. On the same line as the patent number, definite symbols are thus recorded at definite places to indicate the aspects presented by the patent. With such a list it is possible to delegate to administrative personnel the task of retrieving patents corresponding to a given set of aspects; this task involves working through the lists to identify the relevant patents and removing the latter from the collection and delivering to the examiner the combination group that has thus been temporarily constituted. It seems to me that marginal selection card systems, Batten systems, and punched-card systems, in all of which each perforation has a unique significance, can offer little advantage in principle over such systems based on lists. On the other hand, such mechanical systems enable the search by the administrative personnel to be conducted more rapidly and with greater accuracy. In addition, marginal selection systems offer the possibility of recording summaries of the document on the cards themselves, while in the Batten system the choice of the aspects is easier to modify.

Finally mention will be made of a card index that is in use for the azo dyestuffs. To start with, two cards are prepared for each substance, the index being duplicate; for each substance a card is prepared corresponding to the diazo compound and another card is prepared for the coupling component.

The structural formulae are represented by a code. The code consists of two parts, one for the carbon skeleton and the other for the substituents. In both cases a relatively simple code is used, which does not express the structural formula in every detail. This is done deliberately, since we are here concerned not with allotting a unique code to each structural formula, but with the construction of a system which permits rapid retrieval of documents pertinent to structural formulae of a given type. The code for a structural formula is thus composed, as it were, of code units in such a way that the units remain recognizable. Description of a simple example will give some idea of the code system:

chlorophenolsulphonic acid 

is alloted the code C 11: 10–4–610, in which the units have the following significance:

C	carbocyclic compounds
C 1	ditto, with carbon skeleton composed of rings that are not connected by carbon chains
C 11	benzene
:	precedes the code for the substituents
10	—OH (1=singly bound oxygen; 0=hydrogen)
–	separates the substituents
4	halogen
610	—SO_2—OH (6=—SO_2; 1=singly bound oxygen; 0=hydrogen)

A definite order of precedence is laid down for the various components of the code, and the cards are arranged in this order in such a way that a given compound comes in a definite place and can be looked up directly. The cards are interspersed by guide cards designed according to a carefully chosen system. In the example above, the corresponding cards are located behind the fourth guide card governing benzene derivatives. Thus no. 1 refers to code C 11, no. 2 to 10, no. 3 to 4, and no. 4 to 61. The guide cards make it possible to look up any desired combination of the individual code units, either alone or in combination with other unspecified code units. To this end the sequence of guide cards is determined behind the last of which the desired combination would be situated, and one then proceeds to locate the guide cards concerned. For example, if one has to look up a phenolsulphonic acid, the code for which is C 11: 10–61, one encounters the relevant cards behind the guide cards

C 11, 10, and 61. Having discovered guide card C 11, and the existence of a guide card 10 for the first substituent, one looks to see if guide cards subsidiary to the latter are in position for a second substituent (the protruding tongues of the guide cards are situated at different positions along the top edge, and cards bearing a code for a first substituent are arranged with their tongues in line while cards designating a second substituent are chosen with tongues slightly displaced relative to those of cards denoting a first substituent and so on). One inspects successively the guide cards for a second substituent subordinate to 10 until one reaches guide card 61, behind which will be found cards relating to phenolsulphonic acid. Subordinate to this guide card may be found guide cards for a third substituent of lower hierarchy than 6 (e.g., 7, 8, or 9). If a third substituent is of higher hierarchy, e.g., 4 (halogen), the substance (halogenophenolsulphonic acid) is found further forward in the system, namely behind third substituent guide card 61, subordinate to second substituent guide card 4, subordinate in turn to first substituent guide card 10. If one is interested in a phenolsulphonic acid in general, irrespective of other substituents, a series of positions in the system will thus have to be inspected, but within a relatively short range.

The system embraces 6,000 patents, for which some 30,000 cards have been made; the search instructions can be formulated very precisely, and the search can occupy anything from a few minutes to at most, an hour. A disadvantage of this system is that patents in which many substances are mentioned require a considerable expenditure of time for coding, and involve a large number of cards for their incorporation in the system. For the rest, it is a system that meets the most stringent requirements. A characteristic (structural formula) is coded in terms of units in such a way that a search can be instituted for any desired combination of units. The use of one card for each characteristic guarantees that the various units into which a characteristic is broken down are not simply all mentioned somewhere in the patent recorded on the card, but that they are mentioned in combination as that characteristic.

SPECIAL FEATURES OF THE SEARCH FOR NOVELTY IN PATENT SPECIFICATIONS

In order to obtain a good understanding of the reason that led to the system described subsequently in this paper, it is necessary to consider for a moment more closely the special character of the search for novelty in patent specifications. This special character is a consequence of the typical features of patent specifications as well as the search for novelty.

THE PATENT SPECIFICATION

A patent specifies a right granted by a state to a person or corporate body, based upon the patent law of that state. It is therefore a document with a juridical character.

Usually a patent law requires the applicant to describe and explain his invention in an application for a patent in such a way that his description enables a person skilled in the relevant art to reproduce the invention. This description is then incorporated in the specification of the patent when it is granted. In many cases the applicant considers this requirement to be contrary to his own interests and the descriptions will therefore often be at the borderline of what is allowable in this respect.

The juridical character of the patent specification results in the use of unusual terms in describing the invention, so as to cover all foreseeable alternative technical possibilities, and also the contention that the invention is useful for many applications (that usually have not been tried); in short, in a very broad and technically not very detailed picture of the invention. All this is accentuated by virtue of the fact that, first, in the interest of prompt filing of the applications the inventions are described at an early stage of their development when often a good terminology in the art is not yet available, and, secondly, the specifications are drafted by patent experts, each of whom has to cover a wide technical field and consequently often has difficulties with applications in special technical fields in which he is not technically expert.

On the other hand, since patent laws usually require applications for patent to be restricted to a single subject of invention, patent specifications have the character of monographs, which is advantageous from the point of view of classification.

To summarize, a patent specification is thus mainly a monograph in which the accent lies upon something new that is described, in which a fairly sharp formulation of the exclusive rights claimed is contained, but of which the contents, from a technical point of view, are not always clear and in many cases not completely justified.

THE SEARCH FOR NOVELTY

When an examiner examines an application for its patentability, he finds in it a combination of characteristics which the applicant considers in the first place to be new and in the second place so different from what is already known that the normal person skilled in the relevant art would not automatically hit upon this combination. These are two of the most important requirements for patentability in most countries having an examination proced ure. The examiner must verify that these requirements are fulfilled, and will for that purpose among other things conduct a search for novelty in patent specifications. If it emerges that the combination of characteristics sought has already been described in a single patent specification (that was published before the filing date of the application), then this constitutes a bar for granting another patent on this combination. In that case the examiner has no interest in finding any other specifications. If no patent specification is found containing the full combination of characteristics sought, the examiner has to collect sufficient material to obtain a clear picture of the state of the art, in order to be able to assess the level of invention of the application. He will be able to do this if he selects some less restricted combinations of characteristics that still give a good approximation to the invention, and then studies all patent specifications found for these combinations. It is here the desire to locate all specifications describing the selected combinations, just as with literature searches for other purposes.

As the foremost feature of mechanical search for novelty I see the very great importance of the possibility of making a very sharp selection, combined with the possibility of widening the search in very small steps (4). Next it seems to me that also the nature of the combinations of characteristics searched for strongly bears the mark of the search for novelty, and there are no restrictions with regard to applicability in practice. This also influences the nature of a selection system.

REQUIREMENTS TO BE FULFILLED IN A VERY HIGH DEGREE BY A MECHANICAL SELECTION SYSTEM FOR SEARCH FOR NOVELTY IN PATENT SPECIFICATIONS

First, it must be possible to locate relevant patent specifications irrespective of unclear or unusual terminology in their text. This requirement can be fulfilled in a much higher degree by a system based on selection of *notions* than by a system based on selection of words (as, for example, in a Uniterm system).

Secondly, because in formulating inventions notions are used that are as wide as possible, and because each generic notion searched for (in combination with other notions) is fully present in all its species, it is extremely important that when searching for a wide notion all its species are also found. Therefore the use of a code that expresses the essential hierarchy is very important (3, 5).

In systems in which this is not possible (Uniterm system), either, when formulating search criteria, all possible species of the notions to be searched for must be taken into account or, when analysing and indexing specifications for each characteristic contained, all its generic characteristics must also be brought into the system. In both cases this is quite a problem.

A mechanical system will only justify the work put into it and the inconvenience caused by the inevitable additional steps between the formulation of a search order and the study of the specifications found, if a very sharp selection on restricted subjects and a wide variation in search orders is possible (see, in this respect, “The search for novelty” above).

THE LIMITS OF THE FIELD OF TECHNOLOGY COVERED BY THE SUBJECT OF THE INVESTIGATION “CARBURETTORS”

The subject of investigation is limited to what could be called the more or less ordinary carburettors for internal combustion engines, since for this type only is there a need for a possibility of extremely strict selection. Even this limited subject embraces about 16,000 patents!

The subject is defined as follows:

Apparatus for forming and making suitable for use in an internal combustion motor of a combustible mixture of: *(a)* air sucked in by the motor; *(b)* a liquid fuel or liquid fuels supplied by means of the subatmospheric pressure brought about by the motor, which fuel is mixed with the air by spraying (atomizing) it; *(c)* any added constituent.

Fuel supply systems of the level control device of a carburettor are not dealt with as a part of the carburettor.

INTRODUCTION TO THE SCHEME OF CHARACTERISTICS

First, this scheme is based on the point of view that very many objects, and in any event carburettors, can be defined by specifying: *(a)* characteristics which describe a functional part, constructional element, etc.; *(b)* characteristics which indicate the relation between such parts, elements, etc.

In principle it is possible to define a machine, an apparatus, a construction, a structural (chemical) formula (3), and so on, by summation of their functional parts and indication of the relations between them. How far such a definition of objects can be realized in practice is of course a matter requiring separate study in each case. Furthermore, we took the point of view that in cases where it is important to be able to express variant embodiments of the characteristics, this must be effected in such a way that expression is also given in the code to the genus characteristic under which a species characteristic falls.

If the above-mentioned conditions are fulfilled, the code for a given species of machine, apparatus, etc., includes the codes of all genus forms under which the species falls.

Secondly, one can consider carburettors to be characterized by flow paths for various fluids (fuel, medium for the combustion of same, e.g., air, supplements, and mixtures of these), by regions of confluence or divergence of the flow paths (referred to henceforth as flow-path junctions), by devices situated in these flow paths (restricting devices, heating devices, mixing devices, etc.), by the means adapted for the control of the various functional elements, and by further relations between the fluid flow and the devices.

The flow paths are characterized completely by the fluids for which they are intended. The flow-path junctions constitute a form of relation between the flow paths.

The devices derive their raison d'être from the function they are intended to perform. Sometimes it is important merely that a device be present in order to perform a certain function; sometimes it is precisely the constructional details of a device that are of special importance. What is necessary, then, is characterization of the devices according to their function and, in so far as details that could be important are given, characterization of details. Now it is possible in many instances to ascribe more than one function even to a simple device. Thus a carburettor jet serves to form a mixture, but it can also serve to introduce extra fuel for starting purposes. However, the essential function of a jet as such, at any rate in a carburettor, with which this study is concerned, is to form a mixture, and the function during starting is rather an incidental use; the jet can be put to such a use, but it can equally well be made to perform other incidental functions. By characterization according to function is to be understood the essential function of the object as such in a carburettor (the characterization of incidental functions is also important, but this is kept separate from that of the device).

Thirdly, with a view to coding, a system of 15 numbers is made use of for the scheme of characteristics. In order to be able to represent each number of the system by a single symbol, the numbers 10 to 15 are represented by letters, as follows:

10=t; 11=a; 12=b; 13=c; 14=d; 15=e.

Fourthly, the scheme of characteristics comprises four sections:

A Main aspect
B Specification of detail of the devices
C Constituents of the flow
D Relationships; Control; Incidental functions

A characteristic is represented by a single code symbol for the main aspect, by up to four code symbols for specification of the details, and by one code symbol for each flow constituent. In all cases only one main aspect is indicated and one aspect of detail, but all important constituents are indicated. To indicate an alternative aspect of detail the entire characteristic is repeated.

A relationship, control, or incidental function is indicated by one code symbol for the main aspect, and up to four code symbols for the details. In addition, indication is given of the rows on which are recorded the characteristics to which the relationship, control, or incidental function applies. The last indication thus connects with the relationship, control, or incidental function, the characteristics with which the said relationship, etc., is associated. This connection will henceforth be termed the interrelation. We are of the opinion that the manner of coding indicated above only acquires practical significance for mechanized novelty search by the introduction of interrelation symbols. We have noticed in the meantime that such symbols are already in use by the Research group of the U.S. Patent Office (2, 5, 6). This group has made use of them for novelty searches relating to chemical substances. Such interrelation symbols are considered in greater detail under "Perforation of the interrelationships," in the sections on "Punched card and machine" and "Interfixing."

CODING SCHEME

A Main aspect

1.	Supply devices with special means for mixing one flow into another (e.g., fuel nozzles)
2.	Shutdown or restriction devices, e.g., throttles (for float chambers see A=3)
3.	Level control mechanisms
4.	Transport devices
5.	Heating devices (heaters)
6.	Cooling devices
7.	Devices for the improvement of the mixing
8.	Strainers
9.	Devices outside the carburettor; measuring devices
t.	Mixing devices
b.	Devices to split the constituents of a flow
c.	Carburettor
d.	Relations between more than one device and/or flow paths; control systems and incidental functions of devices
e.	Other devices or functions

B Details of devices

A=1 *Supply devices with special means for mixing one flow into another* (e.g., fuel nozzle)

B=1	With spraying devices
11	. with a spray jet
1e	. other devices
e	according to another principle

A=2 *Shutdown or restriction devices, e.g., throttles* (for the float chamber see A=3; supply devices with spray nozzles A=1, B=1)

B=1	With continuously controllable passage
11	. Working successively on a number of passages
12	. and formed as a non-return valve
2	With alternatively a full open or shut-off passage
22	. and formed as a non-return valve
3	With a non-variable restricted passage adjustable or not, including devices in which a fluid jet is either adjustably or non-adjustably directed towards an opening (If there is a controlling flow, this has to be coded, too.)
31	. By means of restriction in a pipe line (metering jet)
7	Adjustable stop block for throttle devices which stop block does not directly affect the flow.
8	Counter devices, other than those mentioned under .. 5 or .. 6
9	Damper for throttle devices
e	Other devices
	Concerning all the shut down or restriction devices mentioned under B=1 ... or 2 ...
..1	.. with an axis perpendicular to the direction of the flow, except a cock (e.g., a butterfly valve)
.. 2	.. movable in the direction of the flow (e.g., spindle valve— needle valve—resistance body)
.. 21	…ball ppvalve
.. 3	.. Slide valve
.. 4	.. Cock (tap)
.. 7	.. Disk slide valve with a flat disk rotatable around an axis parallel to the direction of the flow
.. e	.. other details (the subdivision of A=3 to A=c is omitted)

C=constituent of a flow

1	air
2	fuel
2+3	the lighter of two fuels which are used
2+4	the heavier of two fuels which are used
5	oxygen
6	ozone
7	water or steam

8	lubricating oil or the vapour of it
9	acetylene
10	hydrogen
11	fuel vapour
12	exhaust gases
16	other constituents

D (*A*=d) *Relationships; Control systems; Incidental functions of devices*

B=1	Relative position of devices or flow paths (subdivision is omitted)
2	Coupling between devices (whereby the devices may be governed separately or together)
21	. Mechanical coupling
22	. Coupling with a liquid linkage
24	. Coupling with a gas linkage
28	. Electric coupling
	A combination 1, 2, 4, and 8 can be coded by addition of the codes of all members of the combination, e.g., 4+8 (=12) =b.
	For all relations B=2 is available:
2.1	Devices are built together; two functions in one device
3	Flow paths and/or devices which characterize one of a number of carburettors in parallel connection and linked together and of which the code is the same
3n	. n carburettors
4	Junctions of flow paths
41	. Where 2 or more flows come together in the junction
411	.. the main spray jet lies in the junction (subdivision is omitted)
42	. where 2 or more flows diverge from the junction
5	Throttle device combined with counter throttle device or stop-block or damper
62	. Devices in one flow path
7	Other relations
71	. Devices and junctions coinciding at one spot
Control system and incidental functions of devices	
9	Control system
91	. Controlled by foot or hand
911	.. for adjustment only
92	. controlled in dependence on the pressure or the velocity of a medium
921	.. of the atmosphere
922	.. of cooling water of the motor
923	.. of exhaust gases in the exhaust pipe of the motor
924	.. of the gases in the suction pipe of the motor (926 takes precedence)
925	.. of the medium in the lubricating oil pipe (926 takes precedence)

926	.. of the controlled medium itself
92e	.. dependence on another pressure or velocity
93	. controlled in dependence of a temperature (subdivision is omitted)
94	. controlled in dependence of the viscosity of a medium
95	. controlled in dependence of the r. p.m. of a motor (subdivision is omitted)
9e	. controlled in dependence of other circumstances (subdivision is omitted)
t	incidental functions
t1	. of devices only (i.e., not of the flow in which devices are placed)
t11..	starting, idling or compensating
t111	... starting
t112	... idling
t113	... compensating
t12	.. preventing difficulties resulting from abnormal position of the carburettor
t13	.. preventing backfire
t14	.. preventing gasification in the suction line when motor is stopped
t15	.. collecting and after that remixing of condensed medium
t1e	.. otherwise
t2	. for a flow path (with inclusion of the devices in it) (subdivision is omitted)

PUNCHED CARD AND MACHINE

ONE CHARACTERISTIC PER ROW OF THE CARD

Here the proposals of Luhn (8) and Samain (9) have been followed, in which a code for a characteristic, consisting of a series of letters and/or figures, is not punched in the usual way in a multi-column field of a standard 80-column punched card (one column for each letter or figure), but in one single row; thus in each row may be punched one characteristic. Each row will show a pattern of holes representing one complete characteristic. The selection machine must therefore be able to sense and assess each row separately. Twelve characteristics and/or relations may be punched in the twelve rows of an otherwise standard type punched card. Standard sensing mechanisms of punched card machines sense the rows of the cards consecutively. Therefore the same set of sensing brushes will sense consecutively up to twelve characteristics, which can be recorded on one card, one on each row. The machine, however, must be able to sense each row irrespective of the perforations in the preceding rows; commercially available machines are not able to do this.

PERFORATION OF THE INTERRELATIONSHIPS (INTERFIXING)

Furthermore, the machine must be able to verify that certain characteristics and a relation belong together. If two characteristics P and Q, interrelated by relation R, are sought, the machine would yield up too many cards if it were to select simply all cards in which it found the patterns for P, Q and R, since the relation R would not necessarily pertain to the characteristics P and Q, even though both of these were present, but might equally well belong, for example, to P and Y, or Y and Z, where Y and Z indicate two other characteristics that happen also to be present. It is therefore necessary to indicate and for the machine to verify that the relation R indeed pertains to the characteristics P and Q, in other words, that there exists an interrelation between P, Q and R.

The difficulty here is, that P, Q, and R may each appear on any row of the card because without serious wastage of space it is impossible to assign to each characteristic or relation a predetermined location or even an uninterrupted sequence of locations on the card. However, it is possible to punch in such an order that for each interrelation the machine finds the relation always after (or always before) the characteristics to which it belongs. As the cards are sensed from the bottom row upwards, this means that in each card the relations have to be punched in rows above those used for the characteristics. We envisage the following manner for coding these interrelations.

In each row in which a relation is punched, perforations are also made to indicate the rows in which are punched the characteristics to which this relation pertains. For this purpose the punch positions of twelve columns are invariably used, each corresponding to one row of the card. Therefore if characteristics P and Q are recorded in rows 8 and 9, the row for the relation R, in addition to being punched in a manner indicative of the nature of the relation, is punched in the two columns of this interfixing field corresponding to the rows 8 and 9; if, however, characteristics P and Q are recorded in rows 3 and 7, the row for the relation R is punched in the columns of the interfixing field corresponding to rows 3 and 7. The selection machine must now be provided with a memory for each characteristic sought, that remembers in which row(s) this characteristic has been encountered. In searching for two characteristics P and Q and a relation R between these two characteristics, comparing circuits are set up for each characteristic and for the relation. The memories remember during one card cycle on which row(s) each of the characteristics sought has been found. The comparing circuit for the relation verifies the character of the relation as well as the interfixing, i.e., in the latter respect sees whether in the interfixing field at least those columns are perforated that correspond to rows in which the characteristics were found.

As far as we know there are no machines commercially available that are able to conduct this method of selection. However, there are good grounds for believing that several commercially available sorting or collating machines can be adapted to our purposes without very great difficulty. In any case it appears that the Ilas Scanner (6) of the U.S. Patent Office completely fulfils the above requirements. Accordingly, it does not seem appropriate to go any further into the problem of the machines. It only remains to be mentioned that the development of the machines required is at present in a stage of preliminary discussion with the manufacturers. An initial test of the coding system can be done by using a standard collating machine, with a separate card for each characteristic. Such a test, covering about 1,000 patent specifications will probably be carried out a few months from now.

THE SUBDIVISION OF THE CARD

columns	1–8	spare
"	9	used for a perforation indicating the country in which the patent has been issued (punched in the usual way)
"	10–16	punched for indicating the patent number (usual way)
"	17–20	spare
"	21–40	punched row by row using 1–2–4–8 coding
"	21–24	main aspect
"	25–40	details of devices; nature of relationships, control and incidental functions
"	41–44	spare
"	45–60	punched row by row for indicating constituents; however, one column has been assigned to each constituent, in order to make it possible to record more than one constituent in each row
"	48–59	these also serve as interfixing field, this being possible because these interfixings are only punched in the same row as a relation, control or incidental function, and these are not coded for constituents. Here column 48 corresponds to row 12, column 49 to row 11, column 50 to row 0, column 51 to row 1, etc.
"	61–79	spare
"	80	coded for indicating additional cards A perforation in position 12 indicates that one or more succeeding cards refer to the same patent specification. A perforation in position 1 indicates the first of the additional cards, in position 2 the second, etc.

THE CODE 1–2–4–8 AS APPLIED

For ease of understanding, a table is appended showing this code, that is punched in four columns of one row; the perforations are indicated by a cross (x):

	1	2	4	8
0				
1	x			
2		x		
3	x	x		
4			x	
5	x		x	
6		x	x	
7	x	x	x	
8				x
9	x			x
t=10		x		x
a=11	x	x		x
b=12			x	x
c=13	x		x	x
d=14		x	x	x
e=15	x	x	x	x

EXAMPLES

EXAMPLE I: GERMAN PATENT SPECIFICATION 833,883 (FIG. 1)

In this specification a carburettor is described with a throttle valve in the form of a slide valve 3, positioned in the mixture flow path above the spraying device 7 and movable by means of a Bowden cable 4 (the air inlet of the carburettor is indicated at 1, the mixture discharge at 2). A control needle 6, reaching into the mouth of the fuel jet is joined to the slide valve 3. The air supply can be regulated by means of a hand-controlled choke 8, which is slidable in the throttle valve. The notch 12 on the choke cooperates with the throttle valve 3 in order to limit the lowermost position of the choke. This position lies more to the fully open position as the throttle valve is drawn upwards.

This carburettor has thus an air flow path and a fuel flow path converging at a junction where a flow path originates for a mixture of air and fuel. The main jet lies in the junction. In the air flow path lies a continuously controllable slide valve, in the fuel flow path a continuously controllable needle valve and in the mixture (air with fuel) flow path a continuously controllable slide valve. The valves are built together. The needle valve is joined to the throttle valve so that these are coupled mechanically. By means of the notch 12 the throttle valve and the choke are also coupled mechanically.

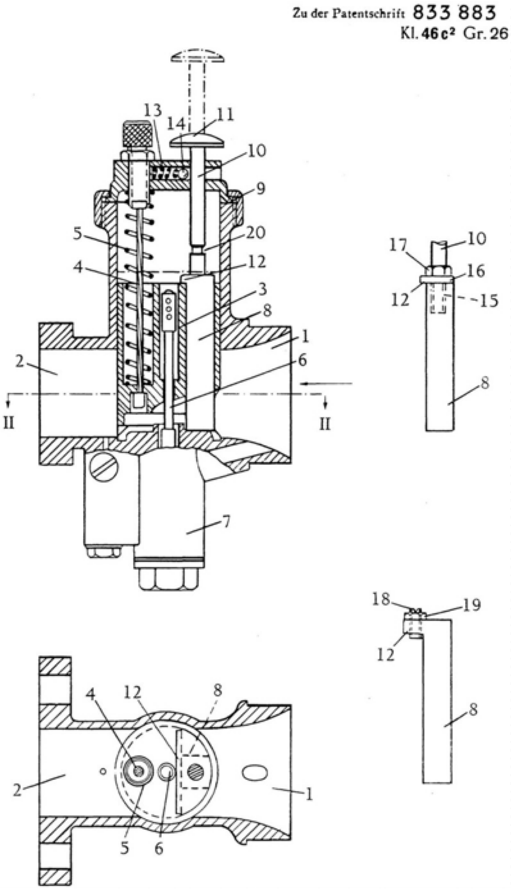

Zu der Patentschrift 833 883
Kl. 46c² Gr. 26
13
14
11
10
9
5
20
4
12
3
8
1
2
6
II
II
7
17
10
16
12
15
8
18
19
12
8
4
12
8
2
1
5
6

FIGURE 1

These data are coded on a punched card as follows (Fig. 2).

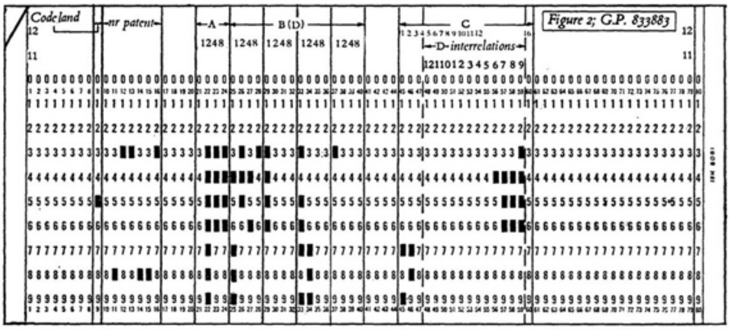


FIGURE 2

Row 9: characteristic I: a continuously controllable slide valve in an air flow path

A=2 (=shutdown or restriction device)

B=1–3–(1=with continuously controllable passage; – –3=slide valve)

C=1 (1=air)

Row 8: characteristic II: a continuously controllable needle valve in a fuel flow path

A=2

B=1–2–(– –2=movable in the direction of the flow, e.g., needle valve)

C=2 (2=fuel)

Row 7: characteristic III: a continuously controllable slide valve in a flow path for a mixture with the constituents air and fuel

A=2

B=1–3–

C=both 1 and 2

Row 6: relation m: the flow paths according to the characteristics I, II, and III have a common junction, in which at least two flows converge; the main jet lies in this junction

A d (d=relation)

B=411–(4=junction of flow paths; 41=.at which two or more flows converge 411=.. the main jet lies in this junction)

interfixing perforations relating to the rows 9, 8, and 7.

Row 5: relation n: the devices according to the characteristics I, II, and III are built together and are also coupled mechanically

A=d

B 211 (2=coupling between devices; 21=. mechanical coupling; 2.1=. devices are built together)

interfixing perforation relating to the rows 9, 8 and 7.

Row 4: relation p: the devices according to the characteristics I, II, and III and the junction according to the relation m coincide at one spot

A=d

B=71– –(71=devices and junctions coincide at one spot)

interfixing perforations relating to the rows 9, 8, 7, and 6.

Row 3: incidental function t: the device according to characteristic I is a choke

$$\left.\begin{array}{l} A = d \\ B = t--- \end{array}\right\} (A = d, B = t = \text{incidental function})$$

B=t111 (t1=incidental function of devices only t11=. starting, idling or compensating; t111=.. starting),

interfixing perforation relating to row 9

EXAMPLE II: GERMAN PATENT SPECIFICATION 678,769 (FIG. 3)

This invention concerns a carburettor in which a preliminary mixture of fuel and air is formed and by means of a spraying device delivered into the suction pipe of the motor. The invention provides means for keeping the preliminary

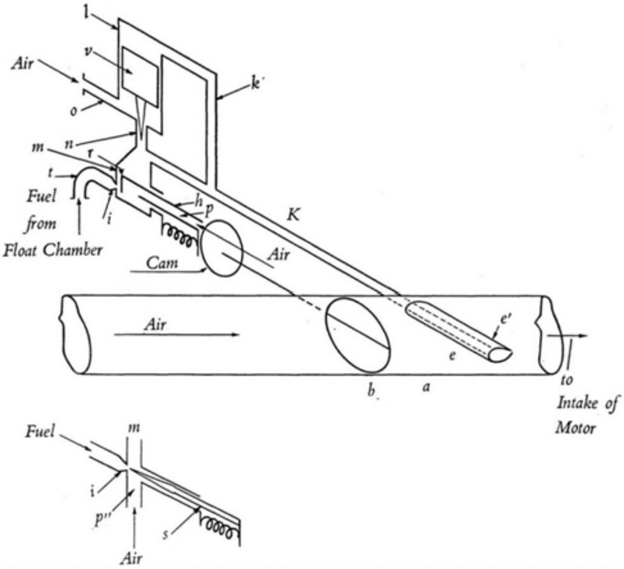


FIGURE 3

mixture under nearly constant pressure, independent of the fluctuation of the subatmospheric pressure in the suction pipe. In this manner the delivery of fuel is independent of the fluctuations of the pressure in the suction pipe. According to the invention there is an additional atmospheric air bleed to the space where the preliminary mixing is effected, which bleed is controlled by a valve dependent on the pressure in the suction pipe, in such a way that the opening of the valve increases with increasing difference in pressure between the atmosphere and the suction pipe.

In the drawing there is a flow path *(a)* for atmospheric air with a butterfly valve *(b)*. The fuel from the float chamber flows through tube *(t)* to the nozzle (i), where the fuel is delivered into the space (m). Opposite the nozzle *(i)* lies the end *(r)* of a slide valve *(h)*. By way of an inlet *(p)* in the slide valve atmospheric air enters the part of space (m) lying under the slide valve. Thence the air flows upwards past the end *(r)* of the slide valve and the fuel nozzle, thereby becoming enriched with fuel to form the preliminary mixture. The distance between the end *(r)* of the slide valve and the opposite wall of the space *(m)* is controllable by movement of the slide valve.

Moreover through a passage *(n)* additional atmospheric air may be sucked into space *(m)*. This air supply is controlled by a needle valve fixed to the piston *(v)*. The underside of the piston is subjected to atmospheric pressure, and its upper side to the pressure in pipe *(k),* which pressure depends on that in the intake pipe of the motor. An increase in the pressure difference between the atmosphere and the intake pipe of the motor results in an upward movement of the piston and thus in enlargement of the passage for additional air. On the other hand, a decrease in the pressure difference results in restriction of the passage for additional air. If the control device is properly proportioned the pressure in space *(m)* will remain nearly constant. The position of slide valve *(h)* is controlled by a cam on the axis of the butterfly valve *(b)*. The slide valve *(h)* can be replaced by a needle valve controlling the opening of the fuel nozzle *(i)*. The piston *(v)* can be replaced by a membrane.

For this example more than 12 characteristics are needed, so it has to be coded on two cards. The interfixing facility is limited to data on one card, so if interfixing perforations are made in a card, this card has to contain the code of all characteristics to which those perforations pertain. For clarity it has been arranged that the first card of the example contains the complete code for the carburettor with slide valve *(h)* in the air inlet, and the second card contains the additional code necessary for recording the alternative construction in which the slide valve in the air inlet is replaced by a needle valve in the fuel inlet. This example (Fig. 4) is coded as in Table 1.

INTERFIXING

METHODS USED FOR INDICATING INTERRELATIONS

In the examples of the preceding section the following relations are discernible between the partial codes or characteristics codes:

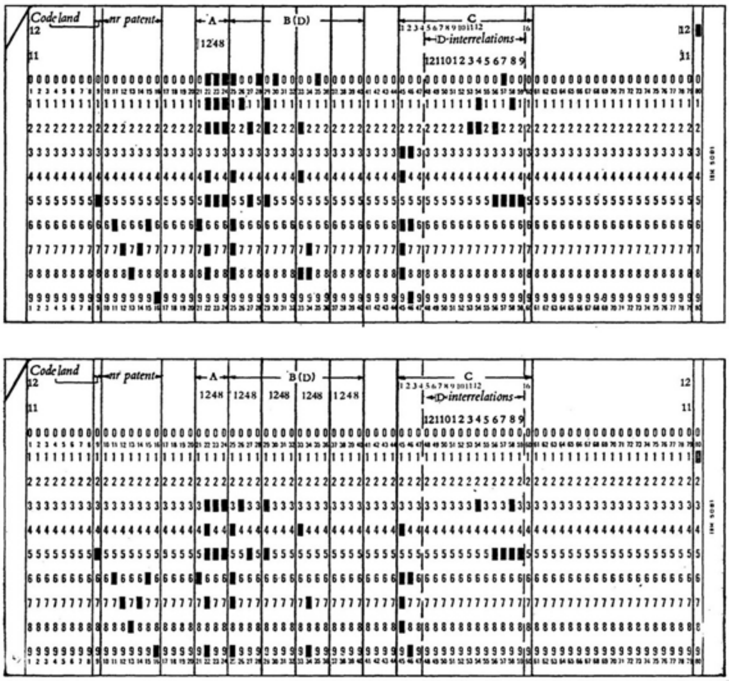


FIGURE 4

1. Between a device coded under aspect B and its genus (as seen in the hierarchy of the code)

2a. Fuel and air both present in the same flow path (as seen by the indication of both constituents in one row, this being made possible by assigning a separate position in the row to each constituent)

2b. Indication for each device of the flow path in which it is located (assigning part of each row for coding one device)

3. The indication of the characteristics to which a relation or control pertains (by interfixing perforations indicating the fields—rows—in which these characteristics have been recorded)

Methods 1 and 2 are often used for recording relations between codes, method 3 seldom.

METHOD 1: CODE SHOWING THE HIERARCHY (3, 5)

This method has the great advantage that in coding a species the genus and their relation is coded at the same time. This happens as it were automatically. If, however, the relation between genus and species is not inherent in the subject of the system, several genuses may have the same series of species. In the latter case it is preferable always to allot the same code to corresponding species; if, furthermore, these are given a fixed location method 2 is obtained.

TABLE 1

Row no.	Characteristic to be coded	Code																			
		A	B				C		Interfixing perforations pertaining to the rows												
							1	2	12	11	0	1	2	3	4	5	6	7	8	9	
	First card																				
9	flow path (t, i) for fuel	.	.	.	.	.		2													
8	a continuously controllable slide valve (h) in an air flow path (p; passage past r)	2	1	.	3	.	1														
7	a continuously controllable valve (needle on piston v) movable in the direction of the flow in an air flow path (o, n)	2	1	.	2	.	1														
6	a flow path (k) for air and fuel ending in a spraying device (e)	1	1	.	.	.	1+2														
5	junction (m) of flow paths mentioned above	d	4	1	.	.											6	7	8	9	
4	air flow path (a) with valve (b) rotatable on an axis perpendicular to flow path axis	2	1	.	1	.	1														
3	flow path from e for air and fuel	.	.	.	.	.	1+2														
2	junction (e) of flow paths mentioned in the characteristics coded in the rows 6, 4, and 3; the main spraying device lies in this junction	d	4	1	1	.								3	4		6				
1	the devices mentioned in the characteristics coded in the rows 8 and 4 are coupled mechanically	d	2	1	.	.									4				8		
0	the valve mentioned in the characteristic coded in row 7 is controlled in dependence on the pressure of the gases in the suction pipe of the motor	d	9	2	4	.												7			
	Second card																				
(12	another card for the same patent follows	a	p	e	r	f	o	r	a	t	i	o	n	in	co	lu	m	n	8	0)	
9	a continuously controllable valve (s, i) in a fuel flow path (t, i)	2	1	.	2	.		2													
8	a flow path (p″) for air	.	.	.	.	.	1														
7	same as for row 7 in the first card	2	1	.	2	.	1														
6	same as for row 6 in the first card	1	1	.	.	.	1+2														
5	junction (m) of flow paths mentioned above	d	4	1	.	.											6	7	8	9	
4	same as for row 4 in the first card	2	1	.	1	.	1														
3	the devices mentioned in the characteristics coded in the rows 8 and 4 are coupled mechanically	d	2	1	.	.									4				8		
(1	this card is a first additional card	a c	p o	e l	r u	f m	o n	r 8	a 0	t)	i	o	n	i	n	r	o	w	1,		

METHOD 2: CODES IN WHICH SUBDIVISIONS ACCORDING TO DIFFERENT POINTS OF VIEW EACH HAVE A PREDETERMINED LOCATION IN A COMMON CODING FIELD

This is a very attractive method, making it possible to build up a large variety of characteristics from a limited number of partial characteristics, in such a way that the characteristics can be inspected from all possible combinations of points of view. If there are many points of view the size of the field needed for coding a characteristic may become so large, that practical difficulties result.

METHOD 3: INDICATION OF INTERRELATIONS BY SPECIAL PERFORATIONS (Compare refs. 5, 6, and 7)

The most important relationships between the (partial) characteristics used are junction relations between flow paths and the locations of devices in a system of flow paths. However, many more relationships exist between devices (e.g., mode of coupling or building together) as well as between flow paths (e.g., their concentricity). It has not proved possible to record them all in a combined code by simple ordering of partial codes. One is therefore obliged either to abandon coding a considerable part of the interrelations existing between the characteristics, or to use special code indications that show which characteristics belong together. The interfixing perforations already mentioned constitute such special code indications. They make it possible, comparatively simply to record much more that can subsequently be retrieved than conventional coding methods. It becomes possible to restrict the codes to characteristics built up to a restricted length from partial characteristics, to record these in an arbitrary sequence and then indicate their interrelations by means of interfixing perforations (in fact this makes the practically usable code length greater by one dimension).

WOULD THE SYSTEM NOT BE SUFFICIENTLY SELECTIVE EVEN WITHOUT INTERFIXINGS?

So far it has always been supposed that interrelations present in the specifications must receive expression. Is this really necessary? Would not it be sufficient to make no attempt to indicate the interrelations but to restrict the coding to those relations that do not receive expression in the relative location of partial codes in a code series? It would still always be possible to find the structures containing the elements of the structure sought. Amongst these several will, of course, not have the interrelations essential to the structure sought, but this would be acceptable if it were only a low percentage. It is better to find too many documents than to overlook some. Are not the possibilities of selection already so sharp that it would still be acceptable to find, for example, twice the number of specifications that would be found making full use of selection on interrelations? Of the numerous theoretically possible interrelations in a code series, are not so few of practical value that leaving out the interrelations would not seriously harm the selectivity of the system?

The final answer to this question can in this case only be given by comparative tests. However, it appears practically certain that the coding scheme for carburettors as explained in this paper would not be satisfactory without the indication of the interrelations. This is partly due to the fact that without the interfixing perforations the difference between many characteristics would disappear. This would be very evident in searching for a combination of characteristics in which the details of the separate component parts were of no importance. Every carburettor has an air flow path, a fuel flow path, a mixture flow path, and a junction of these paths. In the mixture flow path a control mechanism is usually present. Without interfixing, these characteristics are therefore in no way selective.

A search instruction for the invention of Example I of the preceding section without the use of interfixing perforations would practically result in selection of carburettors with a choke and two component parts of any kind that are built together and mechanically coupled. Probably hundreds will meet these requirements. In Example II the selection would practically be only for carburettors having any component part controlled by subatmospheric pressure in the suction pipe. This is also very common. Finally it has been determined that the code of the first variant of the second example would be the same as that of 83 possible variants that are not described in the patent specification.

RESEARCH GROUP

PROGRESS: TIME SPENT

The research group is composed of the following personnel:

The author of this report, for general supervision and guidance (additional job).

Mr. J.Dekker, for code machine problems (additional job). He has specialized in the fields of classification and punched-card machines.

Mr. F.J.Siegers, for guidance during the development of the code and supervision of the coding (additional job). He is an examining officer in the field of internal combustion engines.

Mr. H.van der Esch, as technical assistant, who has also played an important part in the development of the code (virtually full-time job).

Mr. A.A.Looyen, as technical assistant (half-time job).

The development of a usable, unambiguous code has involved: the first three mentioned persons to the extent of 5 months, over-all time; the two technical assistants for 10 months, over-all; an experienced worker can code about 10 patents per day.

REFERENCES

1. Report to the Secretary of Commerce by the Advisory Committee on Application of Machines to Patent Office Operations; Chairman V.Bush; Department of Commerce, Sales and Distribution Division, Commerce Building, Washington 25, D.C. U.S.A., December 22, 1954.
2. A.BLOEMEN. Classificatie in de organische chemie, *Chemische Weekblad,* vol. 37, no. 47 (1940), published by D.E.Centen's Uitgevers Maatschappij, N.V., Amsterdam, pp. 619, 620. Abstracted in *Chemical Abstracts,* “Classification in organic chemistry,” p. 4799, 1942.
3. JAMES W.PERRY, ALLEN KENT, and MADELINE M.BERRY. *Machine Literature Searching,* Interscience Publishers, New York, 1956; especially “Generic encoding of index entries,” p. 12.

U.S.A. Patent Office Research and Development Reports .

4. SIMON M.NEWMAN. Problems in mechanizing the search in examining patent applications 53310-U.S. Dept. of Comm-DC-1956, p. 28, “3. Creating and designing methods of, and apparatus for using, the precise and intricate patterns required.”
5. B.E.LANHAM, J.LEIBOWITZ, and H.R.KOLLER. Advances in mechanization of patent searching, *Chemical Field.* April 11, 1956.
6. Report No. 6., DON D.ANDREWS. Interrelated Logic Accumulating Scanner (Has), June 25, 1957.
7. Report No. 8. Recent Advances in Patent Office Searching: Steroid Compounds and Ilas, 1957. J.FROME, H.R.KOLLER, J.LEIBOWITZ, H.PFEFFER. Steroid Compounds. DON D.ANDREWS. Ilas.

Patents

8. U.S.Patent 2,720,360. Card scanning mechanism, to H. P. LUHN; application date March 31, 1950; patented October 11, 1955.
9. French Patent 994,781, Procédé et machine pour classer et electionner les documents, to M.J.SAMAIN; application date May 15, 1944; patented August 14, 1955. With additions 55,907, 56,342, and 59,504, the last being almost equal to British Patent 662,819.

Descriptive Documentation

CHARLES G.SMITH

One of the advantages of having the proper materials and the ability to use them to make meaningful marks is that a record may be kept of what things are and where they are stored. Where there are many things to be stored, statement in the record of what they are requires distinction between them.

Where the disclosures of documents are concerned, "what things are" becomes a matter of their differing intellectual content. It is the discriminating assessment of this content for the purpose of document identification that poses a problem for which many solutions have been proposed.

An obvious approach to distinguishing between documents is to inquire in what respects some of the documents have disclosures that others do not. From the careful consideration of the thus found characteristics of the disclosures, a list, or a number of lists, of features, none of which applies to all the documents, may be organized upon some stated or readily apparent principle. By keying the list or lists to the location of documents it becomes possible to find documents having a desired disclosure by consulting the list or lists.

Library subject indexes and Patent Office classification schedules are lists of such type. Although they are designed to supply answers to different types of questions, each has the virtue that, if employed as designed, there is a group of documents isolated in which the answer may be found if it exists.

Word or subject indexes as used for libraries have the additional virtue that new subjects or topics relating disclosures to document location may be inserted easily. This is an ability not readily available in the substantially fixed schedules of classification. Patent Office classification schedules, on the other hand, have a certainty of locating documents which concern a specific relation of concepts that is not usually available in an alphabetical index of the library type. It is desirable in mechanized searching for the Patent Office that there be both ease of insertion of new combinations of concepts and certainty of finding of old combinations, if present, in the record searched.

While it is desirable that the valuable abilities of these systems be preserved,

CHARLES G.SMITH Office of Research and Development, United States Patent Office, Department of Commerce, Washington, D.C.

it is also desirable that their disadvantages be avoided in applying them to mechanical search for Patent Office purposes.

The use of word or topic indexing, although suitable for small collections of disclosures centered around some theme, does not appear to have the clarity of approach to specific relations of elements and the sure segregation of combinations of concepts, regardless of how they may have chanced to be expressed in the disclosures, that is essential in Patent Office searches involving many millions of documents from widely diverse technical fields.

For mechanical searching the use of classified lists of possible aspects of subject matter has been proposed. Document identification then becomes a matter of applying to each document a selection from among the multiple aspects available in lists containing graduated groups of possibilities. A problem in doing this is the relation of the lists to one another. Where the variety of arts involved covers the whole of technology and questions requiring specific answers may concern any significant aspect thereof, the number of such lists needed may become unwieldy. Such lists tend also to have the undesirable rigidity now found in Patent Office Classification schedules which renders the introduction of newly discovered aspects of subject matter difficult.

The proposed system is in contrast to such systems as may employ predetermined coding lists expressing various degrees of the probable aspects of documents. Rather it uses, in forming the record to be searched, a manner of organizing concepts such that the more complex concepts may be included in the record as relationships of elementary concepts. The proposed system, for example, does not apply functional names to identify devices or methods but spells out in the record the features, such as the change from one condition to another effected on the material treated, which in their interrelation necessarily characterize them. Any desired aspect can then be read from the record whether it can be named or not. The system thus provides a manner of describing disclosures rather than of classifying them by their aspects.

It is thought that there can be found for each art a group of basic concepts by which persons skilled in the art describe documentary disclosures to each other. Such basic concepts are, in fact, now employed generally in the Patent Office in defining the meaning of classification schedules.

From this viewpoint the problem of documentation becomes a matter of formulating a manner of relating concepts to one another, such that the basic concepts can be fitted together to form more complex concepts and they, in turn, fitted together until whole documents are comprehended. By suitably associating basic concepts, any significant aspect of any disclosure may be recorded and distinctions between documents would rest, as it does now, in the

Patent Office as between classes and between subclasses, upon the relation of the basic concepts applied to one another.

Technical documents are often alike in some respects but differ in others. They, furthermore, differ in the manner in which they include the same idea. In order that the disclosures may be brought to a common basis for comparison, a system of analysis is proposed. The analysis of the document serves as an intermediate stage between the disclosure and the coded record in that it renders the disclosure into the concepts and relations to be used in the coding.

The coding to form the record and the mechanized search thereof must be in the light of the abilities available in some machine. As herein described the system is designed for use with the Interrelated Logic Accumulating Scanner (ILAS) of the U.S. Patent Office (1). Use with other machines is a matter of adaptation to their abilities.

It is thought that by the use of this system of describing disclosures the introduction of a new topic or aspect would be readily accomplished since it would be merely another organization of available basic concepts. Furthermore, since it relies upon the relations of the same ultimate basic concepts as now employed for the purpose, the ability to specify detail as needed is also available.

A DESCRIPTIVE SYSTEM

Since the disclosures of technical documents concern physical things, one way of identifying them in a record is to apply to them various names or terms indicating what things are present in each document. The mere recitation of the terms representing the things present is not enough, however, to discriminate sufficiently between documents where large numbers of documents are concerned. An inventory of the parts present would be of little value in distinguishing disclosures for the type of search required in the Patent Office.

It is the relationships between things that are material in the search for specific devices. Many devices may have substantially the same parts, but they bear little other resemblance to one another. A system to be appropriate for the novelty type of technical search should be able to relate terms to each other so that devices of the same components may be distinguished.

Relationships between terms are not themselves terms and cannot be satisfactorily treated as though they were. A relationship between things is a sort of abstract form into which various terms may be placed. It is necessary, hence, that relationships not only be present in the record but also that the record be clear as to which terms are involved in each relationship.

Where in all the documents for which a record is to be made there is the same predetermined, or inherent relationship between things, it may be feasible to distinguish the documents by naming the things present in the light of such omnipresent relationship. The naming may be made more specific by defining species for each of the things partaking in the basic relationship or combination. Within the disclosures of each patent, however, there are many combinations or relationships that must be in the record.

Where all the documents do not present a fixed relationship between the things disclosed, the system must have some way of including relationships as well as the terms and it must have some way of referring the terms to the particular relationships in which they are involved. The inclusion of a relationship should be made in such a way that various terms may be supplied to it in a uniform manner.

One way (2) of including a relationship in the record and linking terms to it is to apply the relation to each term involved and then connect the terms to one another. Doing this may result in treating relationship as though it were a description of a thing rather than a relation between several things. Furthermore, each thing is involved in many relations to others. It may, in fact, have opposite roles in the same relationship as applied to different other things, and ambiguity may result.

RELATIONSHIP ABSTRACTLY AND BASIC CODING PATTERNS

Another way of including various relationships in the record and specifying the terms to which each applies is to express the relationship apart from the terms and then join terms and relationship to form a combination. This manner of expressing relations is followed in the proposed system.

An advantage of this expedient of the present system is that the ability to establish a chain or sequence of relationship is secured. Such a sequence may be expressed in various ways; for example,

$$a\ R\ b$$
$$b\ R\ c$$
$$c\ R\ d$$

where *a, b, c,* and *d* are conceptually distinct elements and *R* is a relation that exists between them.

The same set may also be written

$$R{:}\ \frac{a\ \ b\ \ c}{b\ \ c\ \ d}$$

or diagrammed

$$R{:}\ a \rightarrow b \rightarrow c \rightarrow d$$

From this basis it is desirable for search purposes that the record be readable

a R d

even though there may be unknown intervening relations as to other elements. Such a result is available through use of the present system.

The coding for retrieval by ILAS which makes this result possible is developed as follows:

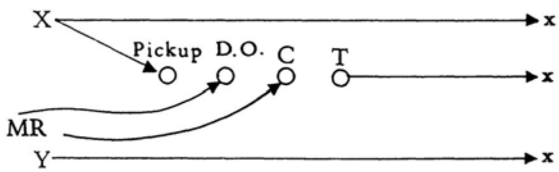
		Interfix
a R b is coded as	a	x
	MR	x
	b	x
	M	

where M is a pulse available for the row in which M occurs.

By the use of a different column in the interfix area for each set the record becomes:

		Interfix area		
a R b b R c c R d } separately encoded as	a	x		
	MR	x		
	b	x	x	
	MR		x	
	c		x	x
	MR			x
	d			x
	M			

The relation, as applied to any two terms X and Y where they are wanted *only if directly related,* may be recovered by this hookup for ILAS:



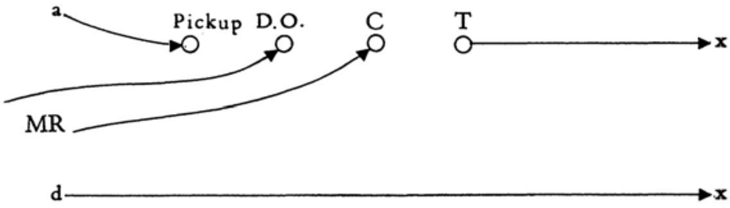
Note that the M pulse applied to the relation is used as a relay opening pulse.

The sets may be related to one another by extending the *relation* interfixing of preceding sets to the columns directed to the other sets:

	Interfix area		
a	x		
MR	x	x	x
b	x	x	
MR		x	x
c		x	x
MR			x
d			x
M			

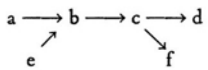
This is the record for a chain relation where a precedes b which precedes c which precedes d.

The a R d relation in the record stated is *indirect*. In order to recover it if wanted whether *direct or indirect* the following hook up is used:



The record is read by ILAS from the top down as shown for the chain relation. Thus with the hookup last shown the finding of “a” will close a relay by pulsing the “pickup.” When the immediately subsequent R appears pulsed by M, a hit is scored in the interfix field and the relay is “dropped out” or opened. This hit will be in all three columns of interfixing pertaining to the topmost R, as shown in the record. Subsequently when the “d” appears and is found interfixed in one of these three columns, a final hit is scored.

A frequent situation which arises in various aspects may be diagrammed:

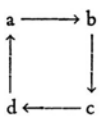


and be recorded as

	a	x				
M	R	x		x	x	x
	e		x			
M	R		x	x	x	x
	b	x	x	x		
M	R			x	x	x
	c			x	x	x
M	R				x	x
	d				x	
	f					x
M						

The same hookup or machine instruction as given previously with reference to the chain of relations is employed for retrieval for this branching chain.

The circuitous situation is also of major importance:



which is recorded as:

	a	x			x
M	R	x	x	x	
	b	x	x		
M	R		x	x	x
	c		x	x	
M	R	x		x	x
	d			x	x
M	R	x	x		x

Here, too, the same machine instructions or hookups used in the retrieval of the chain are applied.

CONCEPTS

The diagrammatic analysis and coding patterns developed thus for expressing relationships are useful in recording a wide variety of situations as found in technical disclosures. They are being applied in this project to methods and machines.

It is, of course, necessary that there be intellectual concepts which lend embodiment to these abstract forms. Another aspect of the present system is the use of a few broad concepts that are widely applicable and commonly present in many arts.

The system reaches for the ultimate concepts which are commonly employed in language and easily understood. These concepts are those necessary to the definition of more specific concepts. A meaning involved in a more specific concept is secured by the use of the power of interrelating broad concepts to each other.

The concepts to be applied to documents may be definable. If they are definable it is in terms of more elementary concepts which must finally have meaning only as based upon actual experience. Thus there is a basic stratum of concepts which need no definition. It is the use of such elementary concepts that is contemplated in the present system. As but a limited number are used and they differ widely, the matter of the choice of which concept to apply is simplified.

An advantage of a documentation system capable of expressing combinational relations is that the elemental concepts employed may be few in number and of a very broad and easily understood nature. This advantage arises in that the distinction between documents would not depend upon the use of a voluminous and highly discriminated terminology, such as is often employed in categorical classification systems, but would rest upon the manner in which simple concepts were organized. The ability to combine concepts implies the ability to express complex ideas by suitable composites of simple ones.

Many words as employed in natural languages represent combinations of more elementary concepts. Such words are a great convenience in spoken or written communication but their use in coding for documentation is undesirable since it places reliance upon such of these words, and their then meaning, as chance to exist at the time of the creation of the codes. Tomorrow may see other words employed for the same composite meanings. Indeed, today, in various arts, different names or words are employed for substantially the same complexes of concepts. Furthermore, there are some combinations of elementary concepts for which no name presently exists perhaps because the circumstances, such as invention, necessary to foster creation of such words have not yet occurred.

The use of summarizing words is unnecessary if the combinations of elementary concepts for which they stand can be expressed directly. It is fundamental in the present system to go behind composite words to the basic organization of elemental concepts which they represent and to set forth the combination essential to the definition of such words.

Where the power of organization is available, complex or composite concepts can be reduced to their elements since the elements may then be organized to express the composite concept.

It may be feasible, as to particular arts, to analyze the facts involved and concepts employed, against the background of a system of possible logical patterns, to arrive at a small group of basic concepts in terms of which all the combinations present in the documents may be expressed. So far as the combinations in them may differ, the documents can be distinctively identifiable, even though the elements and relations per se in them may be the same, since an indicated relation may apply to different elements in each document.

Perhaps a fundamental array of concepts can be found and a common manner of organization employed to the end that all technical literature may be viewed in a common light.

It is to be noted that the present Patent Office Classification Schedules are composed of specific terms which are relations of more elementary concepts. These class schedules are accompanied by definitions setting out the defining combinations of elementary terms. In view of the specificity thus obtained the present use of organizations of elementary concepts is thought well founded.

What concepts in the total that there will be finally cannot be forecast. The list of concepts must grow by the study of particular arts and the carry over from one art to another.

The few concepts necessary to any particular art will be easily grasped and understood, as they are now relative to classification schedules, by those who search the art.

METHOD DISCLOSURES

It is highly desirable where a great number of documents are concerned, as in the present case where there are millions of patents involved, that there be some common foundation or thread running through all of them to which resort may be had to coordinate for various disclosures. If some common basis can be established from which all documents may be approached, then it is feasible to relate in the record each disclosure to all the others.

The linking factor among patents is the method or operation disclosed. All inventions to be patentable must have utility. Utility involves necessarily the use of the invention in some method. Thus all patents concern method.

The primary reason for the existence of machines and devices is the function they perform. Apparatus is created because it has utility in the performance of some method. Apparatus is, in fact, usually denoted by either (1) the activity it causes in some material, such as “agitating,” or (2) the state it produces, such as “vaporizing.”

Furthermore, methods are themselves patentable subject matter. Every class, or nearly every class, in the Patent Office, has, at present, patents directed to methods and usually these method patents are provided for in the class schedule.

It therefore becomes essential to have a concept of method which will furnish a foundation for making a documentary record suitable for patent search. In the proposed system a method is considered to involve a sequence of states and the change from one state to another. The record may thus be formed by expressing the states of a material or of a system of things at successive times and indicating that there is a change in a specified direction from one state to another.

A state may be defined as any condition, status, or activity which can be observed or experimentally determined. It is a sensible aspect of things. It refers to those evidences there are of what a thing is which may be determined by observation and testing. To express it, only ordinary concepts, as used in defining technical viewpoints, are necessary. These concepts are the basic concepts by which complex ideas may be defined.

Method is considered as a sequence of states.

Entity: $S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_4$

R: $Sx \rightarrow Sy$=change of state from Sx to Sy

Sx=state of an entity

Method analysis involves

Successive states that are recognizable if they occur:

a material or object treated or of

a system of entities

The concepts needed to define further a specific process are those also needed to set forth the state or condition of the thing. These will depend on the specific process or art involved. As it is the purpose of this paper to outline a documentation viewpoint and procedure rather than deal comprehensively with the arts, no attempt is now made to state all the concepts of state that may be concerned here.

MACHINE DISCLOSURES

Machines are viewed, functionally, as means for producing a change of state or an effect and for applying energy for that purpose. In producing a change, the operation of the machine is related to the method. It may be said that the machine takes in something in one state and puts it out in another state and that what it takes in includes energy in some form.

Machines have many aspects from which they may be considered. Among them are:

1. Machines of different function may be used in conjunction. There may be a series of devices employed successively to treat the same entity or material with each device making a different change in it. The several machines may or may not take their operational energy from the same source.
2. Machines functioning for the same purpose may employ different structures for doing so. It is desirable that the record be of such a nature that these structural differences may be included in proper relation to the operational aspect with which they are concerned.
3. Even though they function for the same purpose and employ the same structure for doing so, machines may differ from one another by the presence of devices which modify the operation of the machine so that it may better fulfill its purpose. Among such added devices are controls which modify the operation in response to a detected condition or change of conditions. Such devices use energy and such energy may come from the same source as the operating energy, or it may come from some other source.
4. Machines functioning for different purposes may yet have components which are alike or which serve the same subsidiary purpose in both. These components should be identified in the record in similar manner wherever found.
5. Machines must be regarded as both static devices and as operational devices. They may be looked upon as functioning for some purpose as well as being structures the parts of which support and constrain one another. As the latter they must be considered along with other such structures that are not machines in designing a search system.

As it is typical of machines to employ energy, it is thought fitting that as an example of its application to accompany this description of the system an activating means should be employed. That chosen is a hydraulic servo motor system. The choice was guided, in part, by a present need for mechanizing search in that field if feasible. It was also guided by the view that a system suitable for this art would be applicable specifically in other fields such as brakes, presses, pumps, and fluid motors.

OPERATIONAL ANALYSIS

Energy is generally considered to have various forms and to be convertible from one form to another. These forms may be regarded as the states or condition which energy may have or assume and the conversion of energy from one form to another as a change of its state.

A series of changes of the form of energy may be diagrammed analytically as was indicated for a method disclosure above:

Energy: $S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_4$

Energy is not, however, a directly recognizable thing but a concept which is evidenced in the state or condition of some entity. Some of the forms in which energy may be evidenced are the flow of electric current, the flow of a fluid, and the motion of a body in rotation or rectilinearly. In each of these "states" of energy there is some activity of an entity.

Energy in action has the nature that when applied to a device it produces some necessary consequence. When energy is put into a translating device, it produces as a result the output of the device. In the present system this relation is regarded as "with the result that" which is abbreviated as WRT.

Let us suppose we have three devices, D_1, D_2, D_3, each of which receives energy in one form and converts it to another and, further, that these devices act serially so that the output of one device is the input of the next following device.

The devices may be observed to have structural portions which act and co-act as indicated:

D_1 has two portions:

1. A static portion having the connections to which the electric current is supplied.
2. A rotary portion having a shaft which rotates as electricity is supplied.

D_2 also has two portions:

1. A static portion with a chamber having connections to one of which liquid is supplied and from another of which liquid flows as the shaft below is rotated.
2. A rotary portion having a shaft which when rotated has the result that the liquid flows.

D_3 has two portions:

1. A static portion having two chambers separated by a movable wall, a connection to one chamber receiving liquid, a connection to the other chamber discharging liquid.
2. A piston forming the movable wall dividing the chamber of the static portion into two separate chambers and with a projecting rod that moves rectilinearly out of the static portion as the liquid enters one of the chambers.

This disclosure may be analyzed by diagram:

Electric current — Flowing
↓ WRT
Shaft — Rotating
↓ WRT
Shaft — Rotating
↓ WRT
Liquid$_1$ — Flowing — Pressure change: $P_1 \rightarrow P_2$; Place change: reservoir → D_2 chamber → D_3 chamber
↓ WRT
Piston and rod — Moving rectilinearly
↓
Liquid$_2$ — Flowing — Place change: D_3 chamber → reservoir

WRT = With the Result that

In this diagram the vertical arrows point to an activity resulting from another activity. In addition to these resulting activities other features present include:

1. The increase in pressure of the liquid during flow through D_2.
2. The transfer or travel of the liquid from place to place during operation of the device.

These have also been symbolized in the analysis.

CODING OPERATION

The basic coding pattern to be applied to this analysis may be diagrammed:

$$S_1 \longrightarrow S_2 \longrightarrow S_3 \longrightarrow S_4, \quad S_3 \searrow S_5$$

where Sx = an activity of an entity

Sx → Sy = activity x results in activity Y

In accordance with the basic coding the coded record for the devices with regard to energy is:

	Entity	*Activity*		*Area I*				
	Electricity							
		Flowing		x				
M			WRT	x	x	x	x	x
	Shaft							
		Rotating		x	x			
M			WRT		x	x	x	x
	Shaft							
		Rotating			x	x		
M			WRT			x	x	x
	$Liquid_1$							
		Flowing				x	x	x
M			WRT				x	x
	Rod							
		Moving rectilinearly					x	
M								
	$Liquid_2$							
		Flowing						x
M								

Note the placing of the entity involved in a row preceding that in which its activity appears. The only change in hookup required by this for search will be that the entity must trigger a relay to pass the immediately following state of activity of the entity and the relay then be reopened.

The entities considered to be present may be listed as:

	Electricity
D_1	Stator
	Shaft
D_2	Stator chamber
	Shaft
	$Liquid_{1,2}$
D_3	$Stator\ chamber_1$
	Piston and rod
	$Stator\ chamber_2$

These entities may now all be included in the coding:

			Area I				
	Electricity						
		Flowing	x				
		WRT	x	x	x	x	x
D_1	Stator						
	Shaft						
		Rotating	x	x			
		WRT		x	x	x	x
D_2	Stator						
	Shaft						
		Rotating		x	x		
		WRT			x	x	x
	$Liquid_1$						
		Flowing			x	x	x
		WRT				x	x

D_3	Stator chamber$_1$					
	Piston and rod					
	Moving rectilinearly				x	
	Stator chamber$_2$					
	Liquid$_2$					
	Flowing					x

The three devices were indicated to be "transducers" taking energy in and putting energy out in either the same or a different form.

These devices may be bracketed to indicate the input and output states of energy for each. This may be done by applying a like signal for each terminal state. Which is input and which is output is determined by the interfixing. The same signal must be used because the output of one is the input of the one next following.

The record now becomes, using an 8 signal for this purpose:

S			*Area I*				
		Electricity					
8		Flowing	x				
	M	WRT	x	x	x	x	x
		Stator shaft					
8		Rotating	x	x			
	M	WRT		x	x	x	x
		Shaft					
8		Rotating		x	x		
		WRT			x	x	x
	M	Stator					
		Liquid$_1$					
8		Flowing			x	x	x
	M	WRT				x	x
		Stator chamber$_1$					
		Stator chamber$_2$					
		Piston and rod					
8		Moving rectilinearly				x	
	M						
		Liquid$_2$					
8		Flowing					x
	M						

Note in this arrangement that the input state is at the top of the recitation for each device and the output state at the conclusion.

Note, too, that detail of the D_2 stator, a chamber, is on the same line as the stator which it features.

An additional device may be added in the form of a supply chamber, D_4, or "reservoir" for the liquid which goes to D_2 and then to D_3.

With such addition the flow of liquid may be included.

In some devices using this arrangement (servo motors) the liqui$_{d2}$ coming from chamber$_2$ also returns to the "reservoir."

The liquid flow may be diagrammed:
Liquid$_1$ flowing: $D_4 \rightarrow D_2 \rightarrow D_3$
Liquid$_2$ flowing: $D_3 \rightarrow D_4$

The flow of liquid is coded by placing "to" in the row following the source and interfixing both rows with "flowing" and the destination.

Liquid	
Flowing	x
Chamber$_1$	x
To	x
Chamber$_2$	x

Where the flow is to several chambers successively a coding chain appears.

	Area II		
Liquid			
Flowing	x	x	x
Chamber$_1$	x		
To	x	x	x
Chamber$_2$	x	x	
To		x	x
Chamber$_3$		x	x
To			x
Chamber$_4$			x

The flow of liquid may now be included in the coding of the example:

		Interfix areas					
		Area I			*Area II*		
	Shaft						
	Rotating	x					
	WRT	x	x	x			
D_2	Stator chamber				x	x	
	To					x	
	Liquid$_1$				x	x	
	Flowing	x	x	x	x	x	
	WRT		x	x			
	Stator chamber$_1$					x	
D_3	Stator chamber$_2$						x
	To						x
	Piston rod						
	Moving rectilinearly		x				
	Liquid$_2$						x
	Flowing			x			x
D_4	Vessel				x		x
	To				x		

A third aspect of the operation is the change of state of the materials concerned. In the instant device this is represented by the increase in pressure of the liquid by D_2. Such change of state is coded in a third interfix area.

For the pressure change the coding is:

		Area I	*Area III*
D_2	Shaft		
	Rotating	x	
	WRT	x	
	Stator		
	Liquid		x
	Change P_1 Flowing	x	x
	P_2		x

The linking of the interfix areas is by the use of the same row for aspects of the same concept.

CONTROLS ANALYSIS

Devices of a general basic operational design may differ from one another by the controls which are applied to them. Servo motors differ in this respect as to the control of the flow of liquid.

A control is related to the operation of a machine in that some change in the relative position of the parts of the control has the result that there is a change in the operation of the machine in some way. The change of position of a valve, for example, results in a change in the amount of fluid flowing in a system.

In one aspect ("stop-go") a control may have only two states in one of which the operation controlled is permitted and in the other of which the operation is prohibited.

In another aspect of a control device the operation controlled is always permitted but not always at the same rate or in the same degree; the control has the two states of "more" and "less" and change may be in either direction.

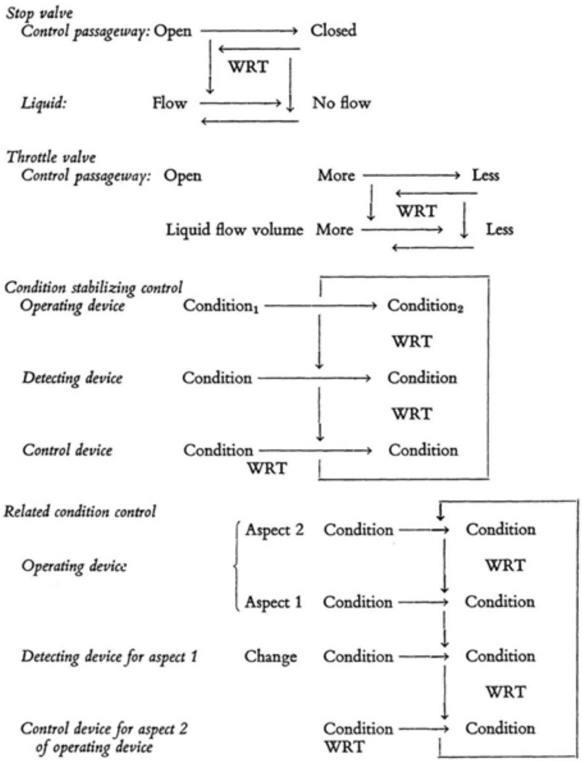
Several controls each of "stop-go" aspect are often employed in conjunction by linking their operators mechanically together. The usual objective of this is the ability to involve different actions either alternatively or alternately. This is done by requiring one control to be "go" when the other is "stop."

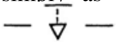
Controls are always operated by the use of energy. The source from which the energy is drawn may be muscular, as in manual control, or it may be of another form. A gravity bias may operate a control; a spring may, too. These energy sources represent energy other than that supplied operationally to the system.

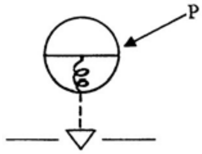
An important source of control operating energy is from the operational system. The system itself may be tapped somewhere so that there is a diversion of a portion of the energy supplied from a source which supplies the operating energy.

Control operation may be made contingent upon some aspect of the operation of the device controlled.

Control of the device by its own internal conditions involves some means which varies with or detects an operating condition of the machine. It also involves a relation of the change which occurs (and which is detected by a change in a detector) to the change made in the control device. Hence with control by machine condition there is a sequence of change relations.



For the analysis of control a simple symbolism may be applied to the basic analytic diagram. In a hydraulic system, for example, a valve may be shown simply as  with an indicator for a hand wheel if manually actuated  or suitable pressure chambers and/or springs if so operated.



Other control devices may be given similar symbolism.

Each of these symbols represents a rather standard coding group of entries based upon the relation of a control to the operation of the device.

CODING OF CONTROLS

In the coding of the flow controls, it is necessary to indicate the relation of the change in the control to the change in the flow activity. It is also necessary to include the energy source which acts with the result that there is a change in the control.

The coding for a manually operated stop valve in operative relation to a flow system is given herewith. Other inputs of additional energy to the system for control purposes, such as bias springs or gravity, would be included as is the manual energy in the example.

	WRT Area I				*Flow Area II*		*Change Area III*	
Liquid					x	x		
Flowing					x	x	x	x
Change			x				x	
Not Flowing							x	x
Change				x				x
Chamber					x			
To					x	x		
Manual								
Action	x	x						
WRT	x	x						
Control Passageway								

Where the control is from an operational change in the flow system a detector is involved which changes as a result of the change in operation of the system. The change in the detector in turn results in a change in a flow controller.

The following coding is for the change in the controller of flow in one line by the change in pressure in another line to the extent that an increase in pressure in the controlling line (liquid$_1$) has the result that flow is permitted in the controlled line (liquid$_2$). This situation arises in servo motors where the application of pressure to an input line to one chamber causes an outlet valve to open in the output line from another chamber.

		WRT Area I			*Flow Area II*		*Change Area III*			
Liquid$_1$ Pressure$_1$							x			
Change		x					x			
WRT		x	x	x						
Pressure$_2$							x			
Position$_1$										
Control Director										
Change		x	x					x		
WRT			x	x						
Position$_2$								x		
Control Passageway										
Open									x	
Closed									x	
Change			x	x					x	
WRT				x						
Liquid$_2$										
Flowing										x
Not flowing										x
Change				x						x
Open					x	x			x	x
To						x				
Change		x	x						x	
WRT			x							
Closed									x	x
Change	x			x						x
WRT				x						
Chamber						x				

SEARCHING THE CODED RECORD

Any of the chains of relations in the various interfix areas may be searched by substantially the same hookups as given previously relative to the basic coding patterns.

There is additionally involved further dependence of one row upon a subsequent row in the same manner as one row was made dependent upon triggering by a prior row in the basic hookups.

Involved, also, is the intersection of cumulative hits registered in plural interfix areas before a final hit is obtained. This requires merely the closing of separate relays by the several cumulative hits with the requirement in the hookup that they all be closed before the final pulse may pass. This is a technique familiar to those who use relay type machines.

It is to be understood that these hookups might be included here but, if they were, they would be with reference to the experimental ILAS machine and of little significance since it has insufficient relay capacity to handle the hookups needed. It is hoped that a production machine will be shortly available.

CONCLUSION

It will be noted that the concepts employed in the analysis and coding are few and of a very broad nature, yet they are so related by the coding that the ability to specify combinations of factors to any desired degree is available

REFERENCES

1. ANDREWS, DON D. “Interrelated Logic Accumulating Scanner (ILAS), *Patent Office Research and Development Report No. 6,* Department of Commerce, Washington 25, D.C., 1957.
2. *Ibid.,* pages 8, 9.

Variable Scope Search System: VS_3

JACOB LEIBOWITZ, JULIUS FROME, and DON D.ANDREWS

This paper will describe a system being developed by the U.S. Patent Office for mechanizing searches for organic chemical compound disclosures. This effort is part of an overall program for mechanization of the entire searching operation of the Patent Office. Its objective is to provide a solution to the complex problems of patent searching occasioned by the rapid, exponential growth of the art, the multiple and variable points of view required in patent search requirements, and the relative inability of rigid manual classification systems to provide such multiple access to subject matter.

The usefulness of the new system is not limited to patent searching. It can be used for any search with respect to chemical compounds in terms of structural characteristics whether the search is done by the patent profession, the research scientist or the industrial organization.

Detailed descriptions of the nature of the patent search problem and some of the mechanization research toward its solution have appeared in the literature (1–11). This paper will give, therefore, only a brief general statement of the problem with respect to chemical compound searching for which the new structural search system is designed.

THE PROBLEM

A patent search is performed with respect to the claimed subject matter of an application for a patent to determine its patentability by comparison with the prior art subject matter. The examiner searches both from the point of view of *(a)* novelty and (b) invention. He is therefore interested not only in identical subject matter but also in similar or the most closely related subject matter. Thus, patent searches are ordinarily from generic points of view regardless of whether the claimed subject matter is for a specific embodiment or a generic class.

JACOB LEIBOWITZ and JULIUS FROME Office of Research and Development, U.S. Patent Office, Department of Commerce, Washington, D.C.

DON D.ANDREWS Director of Research and Development, U. S. Patent Office, Department of Commerce, Washington, D. C.

Another feature of patent searching is the variability in search requirements. The point of view of searching varies in accordance with the constantly shifting emphasis which is characteristic of the field of new developments and invention.

FEATURES OF THE NEW SYSTEM

The search system is called VS_3 (variable scope search system). It is designed to provide multiple access to organic chemical compound disclosures with great variability in scope both generically and specifically.

Coding of the compound for machine searching is done from the structural formula of the compound. The coding is done by nontechnical clerical personnel.

The system permits the use of a simplified coding method. Each compound is regarded as consisting of certain building-block units in certain associations with each other. The building blocks are single ring configurations and selected nonring or chain-unit configurations, e.g., the benzene ring, the diazine ring, and the carboxamide chain unit. The assemblage of codes for these building blocks per se is constant in each compound, while the associations among these building blocks are variable with each compound. The constants, therefore, have preassigned codes and corresponding prepunched cards. In coding a compound, the relationships among these building blocks are indicated and the prepunched cards are assembled and completed according to the indicated associations.

Another feature of the system is the fact that there is no limit imposed on the size of the dictionary providing the descriptive terminology for description of the compounds or the amount of description that can be given to the disclosure of any particular document.

The system is a punched card system using the new machine ILAS previously described (1, 2).

TYPES OF SEARCH QUESTIONS ANSWERED BY VS3

The body of art selected for experimentation with the VS_3 system is the "thiazine" art. The example in Fig. 1 constitutes the disclosure of a compound in this art, taken from U.S. Patent No. 1,996,867.

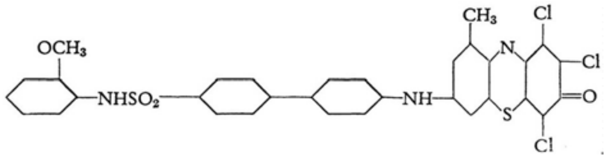


FIGURE 1

The system is designed to permit retrieval of this compound, *inter alia,* on the basis of questions of the following type.

To find:

1. An aryl methyl ether
2. An aryl sulfonamidobenzene compound
3. An azine
4. A thiazine
5. A 1,2,4-trichlorophenthiazine
6. A 1,4-thiazine
7. A dichlorobenzene
8. A sulfonamidodiphenylamine
9. An aminochlorophenothiazine

and so on.

It is important to note that retrieval will be, in each case, not only of the compound depicted but also of any other compounds meeting the requirements of the search. Also, the search varies in scope. It may be expressed in terms of one or several characteristics. The search may be for an ether broadly or a methyl ether more specifically, for a heterocyclic compound, or more specifically an azine, or a thiazine, or a 1,4-thiazine. It can require limitation to positions of substitution or ignore positions of substitution.

VOCABULARY OF THE SYSTEM AND CODES

A standard 80 column IBM card is used for punching the code. Codes are punched horizontally across the card. Thus there may be punched on each card 12 code words each containing 80 bits or punching positions. Figure 2 shows the format of a code word. The first 68 punching positions are divided into 17 characters of 4 punching positions each. The code is punched in hexadecimal digits. Thus this part of the word has available 17 hexadecimal digits.

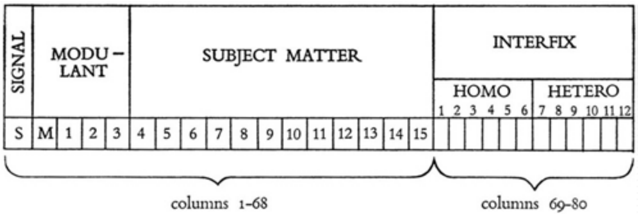


FIGURE 2

This is further subdivided into three fields, a field of one character for signal, a field of 4 characters (M, 1, 2, 3) for modulant and the remaining 12 characters for subject matter. The last 12 punching positions, columns 69–80, is the "interfix" field. This is further subdivided into 6 punching positions for the "homo" field and 6 punching positions for the "hetero" field.

DESCRIPTORS

The terms appearing in the dictionary which are used to describe and define the characteristics of the chemical compounds are designated as descriptors. The descriptors are translatable into code and so the term descriptor and code will be used synonymously.

It is convenient to consider the descriptors and their code equivalents as being of two types, (1) substantive and (2) organizational.

The substantive descriptors describe and define the characteristics of the chemical compounds and they appear in the coding dictionary, a portion of which is illustrated in Appendix A. (The complete dictionary is available at the Office of Research and Development, United States Patent Office). The organizational descriptors express the relationship among these characteristics. This division is purely arbitrary since substantive codes also express relationships. The division is based, however, on the means employed in the system to set forth the relationships.

The substantive codes appear in the 16 characters M through 15 of the word. The organizational codes, expressive of relationships among the substantive codes, appear in the signal and interfix part of the word.

SUBSTANTIVE CODES

The substantive code contains a modulant and subject matter codes. The modulant is a modifier of the subject matter codes; it is a device for using the same code to mean a variety of things according to the particular modulant used. At present, there are 12 modulants employed, although provision has been made for over 65,000 by the allotment of 4 hexadecimal characters to the modulant field. The modulants are listed at the beginning of Appendix A, and the types of information recorded in each type of word according to these modulants are briefly exemplified throughout Appendix A.

ORGANIZATIONAL CODES

There are two types of organization codes, (1) the grouping shown by the signal code and (2) the relationships among the groupings shown by the interfix code.

GROUPING (SIGNALS)

The compound is coded in terms of substructures of various sizes up to and including the complete molecule. The smallest substructural units are the ring and the chain unit. The ring is the familiar cyclic structure; the chain unit is an element or collection of elements in an acyclic configuration. The limits of the chain unit are determined by the groups appearing in the dictionary in the M-1, M-2, M-3, M-4, and M-C words. Most of these groups are the conventional "functional" groups of chemistry although there was no hesitation in synthesizing groups whenever it was deemed necessary from the point of view of retrieval.

The compound of Fig. 1 has been rewritten in Fig. 3 to illustrate the grouping organizations of the structure. The structure has been transformed into the type of skeletal formula set up preparatory to coding. The rings are indicated by Roman numerals; the chain units are encircled. The next substructures to consider are the ring systems and the chains.

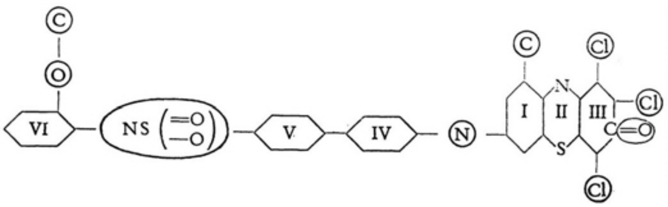


FIGURE 3

RING SYSTEM AND RING

The ring system is a structural entity which comprehends within its scope one ring or a collection of rings in "fused face" relationship.

There are four ring systems in the formula (Fig. 3), as follows:

(a) [(I) (II) (III)]
(b) [(IV)]
(c) [(V)]
(d) [(VI)]

The parentheses are used to symbolize the enclosure of the codes pertaining to a ring, and the external brackets symbolize an enclosure of the collection of rings pertaining to the ring system. It will be noted that a ring system may con tain only one ring as in *(b), (c),* and *(d)*. Rings connected by a bond juncture are not in the same ring system, while rings in fused face relationship are.

A ring system is not only a collection of codes for the individual rings it contains. It consists of other characteristics not present in the individual ring codes, such as the fused face relationship (see Appendix A) and also characteristics of the type provided for by the ring system word M-7 (Appendix A). In setting up the codes for the ring system, then, this additional information is included within the ring system grouping.

CHAIN UNIT AND CHAIN

The relationship between the chain unit and the chain is analogous to the relationship between the ring and the ring system. The chain is a continuity of chain units; its continuity is terminated by the interposition of a ring. It may consist of one chain unit only. The following are the chains found in the compound of Fig. 3 (with the same bracketing symbology).

(a) [(C) (O)]

(b) $[(NS{=O \atop -O})]$

(c) [(N)]

(d) [(Cl)]

(e) [(Cl)]

(f) [(Cl)]

(g) [(=O)]

(h) [(C)]

This is again a substructure within a larger substructure.

COMPOUND GROUPING

The next order of grouping is the enclosure of the whole set as a compound unit. Within this unit codes are then added to indicate additional information pertaining to the entire compound group. The type of information added is indicated in words M-8, M-9, M-A, M-B (see Appendix A).

PATENT GROUPING

The next and final grouping is that which encloses all the compounds as pertaining to the same document. This grouping organization of the codes for a compound structure may be compared to the organization of the written language. The substantive codes constitute the identification of the letters of the alphabet and their meanings in the word; the collection of words constitutes the larger organization of the sentence. In addition, however, since the meaning of the sentence is more than the mere sum of the words, additional substantive information is added pertaining to any additional concepts provided by the sentence as a whole. The sentences then are in a group constituting the larger substructure of the paragraph, with further information pertaining to the paragraph not available as the mere contextual sum of the sentences.

SIGNALS

The grouping of the codes is handled through the signals, a list of which appears in Appendix A. These signals permit the proper correlations among the codes so that the codes for one ring do not get scrambled with the codes for a different ring, or the codes for one compound do not become correlated with the codes for a different compound.

There is no fixed limit to the number of codes that may be included within any signal group. Thus, any number of descriptors may be applied to definition of a ring, a ring system, a chain unit, a chain, or the compound as a whole. By the same token there is no particular limit to the amount of information that can be recorded for any one document.

In the machine operation, the presence of the signal signifies the termination of all the codes pertaining to the particular structural unit defined by the signal.

While properties and functions have not been encoded into the system at present, it will be obvious that this can be done according to the grouping logic (11). Properties of a compound would be grouped on the compound level; properties of a substructure would be grouped within the level of said substructure. Thus in searching, correlations can be made between compound and function or substructure and function.

INTERFIX

Another organizational relationship is that of connectivity of the substructures. In the compound of Fig. 3, ring I is joined to N, ring IV is joined to N and to ring V, ring VI is joined to chains —O—C and —NSO_2—, and so on. This relationship is handled by the interfix device. This involves the assignment of a pair of identical arbitrarily selected numbers to those substructures which are joined to each other.

The following types of connections are defined by interfix:

(a) Ring to ring by bond juncture *(b)* Chain unit to chain unit	Homo interfix
(c) Ring to chain unit *(d)* Ring to chain	Hetero interfix

Types *(a)* and *(b)* are called the homo interfixes and are coded in columns 1 to 6; *(c)* and *(d)* are the hetero interfixes coded in columns 7 to 12.

For illustration, a portion of the formula of Fig. 3 has been extracted and interfix numbers assigned as shown in Fig. 4. Each interfix number has been placed at the bond juncture to indicate that it is assigned to each of the sub-structures involved in the connection. Thus in coding, the units will have the following interfix numbers (interfix in the subscript): (C_1), $(O_{1,7})$, $(VI_{7,8})$, $\left(NS\left(\begin{smallmatrix}=O\\-O\end{smallmatrix}\right)_{8,9}\right)$ (V_9). The substructures of Fig. 4 can be reconstructed according to the rule that those substructures are joined to each other which have the same numbers. It must be emphasized that the value of the number is of no consequence. What is significant is the identity of the pair of numbers.

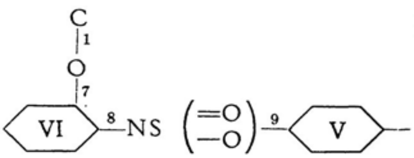


FIGURE 4

Thus, the groups can be renumbered as shown in Fig. 5. The structure is equivalently defined and can be reconstructed from the interfix relationship.

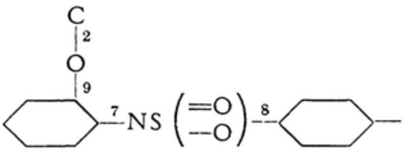


FIGURE 5

The interfix is punched in any of columns 69 to 80 according to the number assigned. While scanning and sensing of the codes occur row by row, determination of the interfix relationship is done by a vertical comparison down the column. Thus one code word in a row is connected to some other code word in a different row, either on the same card or on a different card, through recognition of a pair of punches in the same vertical column. This principle will be further illustrated. See Appendix B.

CODING: SET-SUBSET, EXACT PATTERN

Two methods of recording the codes are used according to two different desired searching devices: (1) the set-subset method and (2) the exact pattern method. To exemplify this, the code words used to describe a ring, i.e., M-5 for set-subset and M-6 for an exact pattern have been selected. In the M-5 word (see Appendix A), assignment of codes is on a bit-by-bit (set-subset) basis. In character 7, for example, the bits have been allotted the meanings shown in Fig. 6. A ring which is both unsaturated and heterocyclic will be coded as 7–5 in the M-5 word as in Fig. 7. A search for a heterocyclic ring 7–4 will involve searching for a hole in the 4 bit and will result in retrieval of the structure. Similarly, a search for an unsaturated ring 7–1 or an unsaturated heterocyclic ring 7–5 will result in retrieval of the indicated ring.

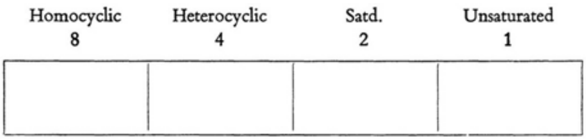


FIGURE 6

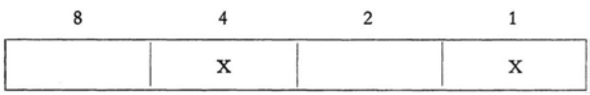
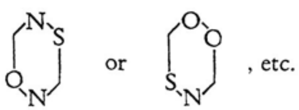


FIGURE 7

Characters 9, 10, 11, 12, 13, and 14 involve the same type of coding. Thus, a search for a ring containing a heterocyclic nitrogen ortho to sulfur will result in retrieval of said structure even though other groups are present such as:



In addition to the bit-by-bit basis, the terms of characters 4, 5, 6, and 8 have been coded by what may be called an "at least" basis, i.e., a ring containing 4 or more nitrogens has at least 3, 2, and 1 nitrogens; a 3 nitrogen ring contains at least 2 and 1 nitrogens, and a 2 nitrogen ring contains at least 1 nitrogen. Thus character 4 of the M-5 word will be punched as follows for a 4N, a 3N, a 2N, and a IN containing ring:

	4N	3N	2N	1N
4 N ring	X	X	X	X
3 N ring		X	X	X
2 N ring			X	X
1 N ring				X

FIGURE 8

The exact pattern coding is illustrated in M-6 (Appendix A). Each code is uniquely different from any other. A 4-membered ring 12–4 is not found within a 6-membered ring 12–6. Similarly, a 5 carbon ring is not found within a 6 carbon ring.

In searching, the codes can be used in any combination desired. Thus, where a search is for a diazine and it is desired to exclude the 3 N-or-over containing rings, the combinations of codes requiring a 6-membered ring, at least 2 nitrogens and exactly 4 carbons will result in this exclusion.

RINGS

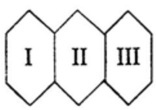
Rings are coded according to the M-5 and M-6 words as already described. In addition to the descriptors found on pages of the coding dictionary in Appendix A, a code called an index number code is assigned in characters 4, 5, and 6 of the M-6 word. A list of rings with their index numbers appears in the coding dictionary. These are unique numbers defining the specific rings as they are disclosed in the *Ring Index* (12). The rings in the list have already been completely coded and are maintained in a card file. In addition, a set of prepunched cards has been prepared which contain the punched codes for each of these rings. When these rings are encountered in a formula, they need not be coded again. They are merely identified and the prepunched cards are selected for further processing. As new rings are found in the disclosures they are coded, unique index numbers are assigned, and they are added to the file.

The signal is coded in a separate word as the last word of the set. Thus, three codes are shown for each ring:

M-5 and M-6 followed by S-1.

RING SYSTEM

The ring system is the sum of the codes for the rings contained within the ring system plus any other codes required to define relationships not provided for by the sum of these codes. Where there is only one ring in the ring system, two codes for the ring plus S-1 are followed by S-2, and the coding for the ring system is complete. Where there are two or more rings in the ring system, additional codes are added to define relationships not expressed by the individual ring codes. One of these relationships is the "fused face" pattern. The fused face pattern defines the relative positions of the fused faces of each ring in the system, in accordance with the graphic portrayal of the coding dictionary in Appendix A. The heavy lines refer to a position of fused face joining with another ring. The patterns are not necessarily limited to carbon rings; the same relationships are intended to be applicable regardless of the kind of ring elements.

Thus the structure 

shows the following patterns for each of the rings:

I pattern A 

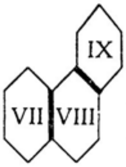
II pattern D 

III pattern A 

The structure

shows the same patterns.

The ring system



shows the following patterns for each ring: VII, pattern A; VIII, pattern C; IX, pattern A. Thus it is possible in searching to discriminate on the basis of ring orientation.

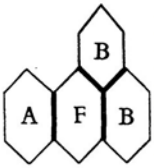
The codes are arranged to provide for finding any pattern contained within another pattern. Thus pattern D is found within F, H, I, J, K, and L but not within A, B, C, E, and G.

A generic search, therefore, on the basis of the following ring system type will

I

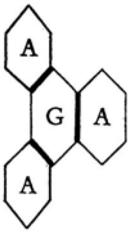
result in retrieval of the following structural type since I is contained within II as follows:

II



I	II
Ring A	is equal to Ring A
Ring A	is contained within Ring B
Ring D	is contained within Ring F

On the other hand, a structure



will not be retrieved since D is not within G.

The ring system is further coded according to the type of information in M-7 (Appendix A). This will provide further variation in scope as to searching. The search for a structure of type I, above, for example, can be made more specific by requiring that there be only 3 rings in the system (code M-7, 7–3), whereupon structure II would be excluded.

Ring systems are also preceded and assigned index numbers in the manner indicated for the ring.

CHAIN UNITS

The term chain unit is one of organization. The terms exemplified in M-1, M-2, M-3, and M-4 of the dictionary are, with certain exceptions, while referred to as chain units, equally useful in describing ring structures. Thus, by inclusion of the code within the ring group, it would become a ring description. For the time being, however, these codes are being used as chain descriptors only.

There are five types of chain units: M-1, M-2, M-3, M-4, and M-C. The M-C word refers to the carbon unit. The M-1 to M-4 types are described by general formulas. Thus, the M-1 group is indicated to be of type ABX (=Y) CD. The letters themselves are of no significance. The formula represents an arbitrary notation system which serves as a guide for writing functional groups in a particular sequence. The groups in M-1 all have an element doubly bonded to another which serves as the focal point with reference to which a sequence is written for the chain word.

While each chain unit has a unique code, there are various subgeneric units found within it. For example, a carboxylamido group M-1, 6–2, 7–3 is found within a urethane, a urea, a semicarbazide, and so on. A search on this basis will result in retrieval of all these groups. On the other hand, the search can be limited to any specific one desired, e.g., carboxyamide per se, M-1, 4–1, 5–1, 6–2, 7–3, 8–5, 9–1, since the code is unique for this unit.

As a further example, (see Appendix A) a search for an amine [M-3, 5–5] will result in retrieval of primary, secondary, tertiary, and quaternary amines, imines, hydrazine, etc. A requirement for a secondary amine, M-3 codes 4–1, 5–5, 6–2, will limit the search to secondary amines only.

These chain units have been generated from a small dictionary of elements, some actual and some synthetic (as “Z”). As more chain units are found, codes for them are generated according to the notation system.

The chain unit codes are terminated by signal S-3.

CHAIN

The S-4 signal groups the chain units. The interfix number of a chain connection to a ring is the sum of the interfix numbers of its chain unit connections to the ring. If a chain is intermediate to a number of rings, i.e., its terminal chain units are interfixed (ring to chain unit), the S-4 for the chain has the interfixes of all the terminal chain units (hetero interfixes). In searching, a connection of a chain or any part of a chain to any particular ring can be found with or without specificity as to the particular chain unit by which it is connected.

CARBON WORD

The carbon word is coded to indicate the number of carbon items both specifically and generically. The specific code permits limitation in searching to a specific alkyl group. The generic code permits the finding of a 2 carbon chain within a 2+ carbon chain, a 3 carbon chain within a 3+ carbon chain, etc. A search can be made for alkyl groups, saturated or unsaturated, straight or branched, lower or higher, as desired.

POSITIONS OF SUBSTITUTION

The fact that the coding of a compound takes place in terms of substructures permits the use of a standard numbering system. Each ring, except for a homocyclic carbon ring, is assigned an arbitrary numbering system and the positions of substitution of chain units are indicated according to this number standard. The same standardization is done for the ring system. The position of substitution of a chain unit to a ring and to a ring system is coded in the chain unit word. The position of substitution of a ring to a ring is coded in the S-1 word of the ring code.

EXAMPLE OF CODING

The coding of the compound of Fig. 1 is illustrated in Appendix B.

SEARCH

The variability in scope of search in this system is:

1. The "building blocks" of the system are small units. These units are separately and independently described, which permits the asking of a large number and large variety of search questions for retrieval.
2. The descriptors are variable in scope—one descriptor may merely indicate a 6-membered ring while another indicates a positional relationship of heterocyclic elements or substituted groups. Search questions may be expressed in terms of combinations of descriptors, which gives variability in scope of the question.

3. Each collection of codes is gathered into a substructural entity. This permits search questions with respect to chemical compounds which vary in scope as desired with respect to selected portions of the molecule.
4. The groupings and interfixes provide the ability to specify relationships among the substructures and to obtain as much specificity as desired with respect to the compound search.
5. The machine used is the ILAS. Scanning of the cards is a continuous operation from card to card, as many cards being used as needed to encode the disclosure of any one document. Termination of the card or cards pertaining to the document is obtained by the signal S-6.

EXPERIMENTAL NATURE OF SYSTEM

The system is being tested. Thus far, it appears to be fully operative according to the principles involved. The coding procedure itself appears to be quite simple. The compounds are coded from the formulas by clerical personnel. The preceding device for the constants of the formulas, rings, ring systems, and chain units has been found to be a very useful procedure from the point of view of speed, accuracy, simplicity of coding, and uniformity.

FUTURE WORK

An effort is being made to solve the problem of the Markush disclosure. The Markush type of disclosure, where a formula is presented with a number of sets of alternatives, has been described (2, 4, 8). It is desired to code this type of disclosure so that in retrieval only one member of each set of alternatives may be selected in combination with only one member of any other set. To handle this problem, a signal is being used to provide a counting method for each group in "and" relationship, so that of a set of alternatives, no more than one is selected.

THE POLYMER SYSTEM

The principles described herein are deemed applicable to a more comprehensive problem such as exists in the disclosure of chemical compositions and processes. By an expansion of the principle of groupings, compositions can be provided for within a higher level of grouping enclosure, and processes can be provided for by a still further order of grouping. This principle is being used in a system now under development for encoding the disclosures in the polymer art which includes compounds, compositions, and processes. The method for describing the compounds in that art is not the same as described herein. However, the VS_3 system is deemed compatible with this more comprehensive system, and it will be for the future to provide a determination as to whether or not the methods should be merged.

ACKNOWLEDGMENT

Acknowledgment is made of the assistance of Mr. H.P.Luhn of IBM in helping develop some of the concepts herein described, particularly with respect to the concept of fused face patterns. Appreciation is also expressed for the assistance of Mrs. R.W.Swanson of the examining staff and Mrs. A.Replogle, Mrs. J.M.Hale, Mrs. M.Bender, and the various other members of the clerical staff in this experiment.

BIBLIOGRAPHY

1. DON D.ANDREWS. *Interrelated Logic Accumulating Scanner (ILAS).* Patent Office Research and Development Report No. 6. Washington, 25, D. C., Department of Commerce, 1957.
2. DON D.ANDREWS, JULIUS FROME, H.R.KOLLER, JACOB LEIBOWITZ, and H. PFEFFER. *Recent Advances in Patent Office Searching, Steroid Compounds and ILAS.* Patent Office Research and Development Report No. 8. Washington 25, D. C., Department of Commerce, 1957.
3. DON D.ANDREWS and SIMON M.NEWMAN. *Storage and Retrieval of Contents of Technical Literature, Nonchemical Information, Preliminary Report.* Patent Office Research and Development Report No. 1. Washington 25, D. C., Department of Commerce, 1956.
4. JULIUS FROME and JACOB LEIBOWITZ. *A Punched Card System for Searching Steroid Compounds.* Patent Office Research and Development Report No. 7. Washington 25, D. C., Department of Commerce, 1957.
5. B.E.LANHAM, J.LEIBOWITZ, and H.R.KOLLER. *Advances in Mechanization of Patent Searching, Chemical Field.* Patent Office Research and Development Report (No. 2). Washington 25, D. C., Department of Commerce, 1956.
6. B.E.LANHAM, J.LEIBOWITZ, H.R.KOLLER, and H.PFEFFER. *Organization of Chemical Disclosures for Mechanized Retrieval.* Patent Office Research and Development Report No. 5. Washington 25, D. C., Department of Commerce, 1957.
7. SIMON M.NEWMAN. "Linguistics and Information Retrieval; Toward a solution of the Patent Office Problem." Monograph Series in Linguistics and Language Studies, No. 10. Washington 25, D. C., Georgetown University Press, 1957.
8. SIMON M.NEWMAN. *Linguistic Problems in Mechanization of Patent Searching.* Patent Office Research and Development Report No. 9. Washington 25, D. C., Department of Commerce, 1957.
9. SIMON M.NEWMAN. *Problems in Mechanizing the Search in Examining Patent*

Applications. Patent Office Research and Development Report No. 3. Washington, 25, D.C., Department of Commerce, 1956.

10. SIMON M.NEWMAN. *Storage and Retrieval of Contents of Technical Literature, Nonchemical Information, First Supplementary Report.* Patent Office Research and Development Report No. 4. Washington 25, D. C., Department of Commerce, 1957.
11. M.F.BAILEY, B.E.LANHAM, and J.LEIBOWITZ. Mechanized searching in the U.S. Patent Office. *Journal of the Patent Office Society, XXXV* (1953), 566–587.
12. AUSTIN M.PATTERSON and LEONARD T.CAPELL. *The Ring Index,* Reinhold Publishing Corp., New York, 1940.

APPENDIX A EXCERPTS FROM VS3 INDEX OF CODES

The codes contained in each word are introduced by a modulant. Twelve modulants are presently defined as follows:

Modulants	M-1 to M-4 for chain units
M-5 and M-6 for rings	
M-7 for ring systems	
M-8 for compounds	
M-9, M-A, M-B for metals	
M-C for carbon chain units	

Signals are employed to group the words pertaining to rings, ring systems, functional groups, and chains as follows:

Signals	S-1 pertaining to individual rings
S-2 pertaining to the ring system	
S-3 pertaining to the chain unit	
S-4 pertaining to the complete chain	
S-5 pertaining to the complete compound	
S-6 pertaining to the complete patent	

Interfixes are employed to define the bond relationships. Twelve interfixes are provided, numbers 1–6 for homo relationships (ring to ring or chain unit to chain unit) and numbers 7–12 for hetero relationships (ring to chain).

MODULANT M-1 *Chain unit* A B X(=Y) C D

Chemical formula	*Name*	*Structural formula*	A	B	X	=Y	C	D	
COF	Carboxylfluoride	C(=O)F	Z 4-1	Z 5-1	C 6-2	=O 7-3	F 8-B	Hal 9-D	
CON	Carboxylamide	C(=O)N	Z 4-1	Z 5-1	C 6-2	=O 7-3	N 8-5	Z 9-1	10-8
CON_2	Isourea	OC(=N)N	Z 4-1	O 5-3	C 6-2	=N 7-5	N 8-5	Z 9-1	10-8
CON_2	Urea	NC(=O)N	Z 4-1	N 5-5	C 6-2	=O 7-3	N 8-5	Z 9-1	10-8
CON_3	Semicarbazide	NC(=O)NN	Z 4-1	N 5-5	C 6-2	=O 7-3	N 8-5	N 9-5	10-8
CO_2N	Urethane	OC(=O)N	Z 4-1	O 5-3	C 6-2	=O 7-3	N 8-5	Z 9-1	10-9
SO_2I	Sulfonyliodide	S(=O, −O)I	Z 4-1	Z 5-1	S 6-4	(=O, −O) 7-8	I 8-9	Hal 9-D	
SO_2N	Sulfonamide	S(=O, −O)N	Z 4-1	Z 5-1	S 6-4	(=O, −O) 7-8	N 8-5	Z 9-1	10-8

MODULANT M-2 *Chain unit* A X (=Y) B

Chemical formula	*Name*	*Structural formula*	A	X	=Y	B
CO	Ketene	=C(=O)	= 4-E	C 5-2	=O 6-3	Z 7-1
CO	Ketone	C(=O)	Z 4-1	C 5-2	=O 6-3	C 7-2
CO	Quinone	C=O	C 4-2	C 5-2	=O 6-3	C 7-2
CS	Thioketene	=C(=S)	= 4-E	C 5-2	=S 6-4	Z 7-1
CS	Thioketone	C(=S)	Z 4-1	C 5-2	=S 6-4	C 7-2

MODULANT M-3 *Chain unit* A X Y

Chemical formula	*Name*	*Structural formula*	A	X	Y	
NH	Secondary amine	>NH	Z 4-1	N 5-5	C 6-2	
N	Tertiary amine	>N-	C 4-2	N 5-5	C 6-2	
O	Ether	-O-	Z 4-1	O 5-3	Z 6-1	10-4
S	Thioether	-S-	Z 4-1	S 5-4	Z 6-1	10-4
N	Quaternary amine	>N<	Z 4-1	N 5-5	Z 6-1	
N	Imine	=N	= 4-E	N 5-5	Z 6-1	
NH_2	Primary amine	H-N-H	Z 4-1	N 5-5	H 6-6	
N_2	Hydrazine	>N-N<	Z 4-1	N 5-5	N 6-5	
Br	Bromine	-Br	Hal 4-D	Br 5-C	Z 6-1	
Cl	Chlorine	-Cl	Hal 4-D	Cl 5-A	Z 6-1	
F	Fluorine	-F	Hal 4-D	F 5-B	Z 6-1	
I	Iodine	-I	Hal 4-D	I 5-9	Z 6-1	

MODULANT M-4 *Chain unit*

Chemical formula	*Name*	*Structural formula*	*Notation code*
CN	Nitrile	−C≡N	4-B
CN	Isonitrile	−N=C	4-6
CN_2	Carbodiimide	N=C=N	4-8
CNO	Cyanate	O−C≡N	4-3
CNO	Isocyanate	N=C=O	4-4

MODULANT M-C *Carbon chain*

No. carbons	*No. carbons specifically*		*No. carbons generically*		CU *Posn. on ring*	*Code*	CU *Posn. on R.S.*	*Code*
	4	5	8	9				
1		5-1		9-1	1	11-1	1	13-1
2		5-2		9-1	2	11-2	2	13-2
3		5-3		9-1	3	11-3	3	13-3
4		5-4		9-3	4	11-4	4	13-4
5		5-5		9-3	5	11-5	5	13-5
6		5-6		9-7	6	11-6	6	13-6
7		5-7		9-7	7	11-7	7	13-7
8		5-8		9-F	8	11-8	8	13-8
9		5-9		9-F	9	11-9	9	13-9

MODULANT M-5 Ring schedule

No.	N *atoms*	O *atoms*	S *atoms*	*Misc. hetero atoms*	*Type of ring*	
1	4–1	5–1	6–1	8–1	Unsaturated	7–1
2	4–3	5–3	6–3	8–3	Saturated	7–2
3	4–7	5–7	6–7	8–7	Heterocyclic	7–4
4,+	4-F	5-F	6-F	8-F	Homocyclic	7–8

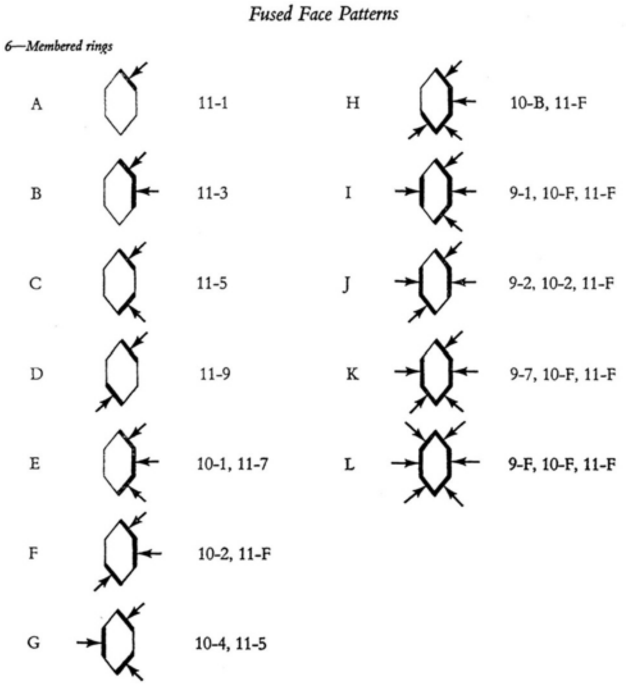
Combinations of atoms	*Ortho*	*Meta*	*Para*
N-N	9–1	11–1	13–1
N-O	9–2	11–2	13–2
N-S	9–4	11–4	13–4
O-O	9–8	11–8	13–8
O-S	10–1	12–1	14–1
S-S	10–2	12–2	14–2
Misc. *(a)*-Misc. *(a)*[a]	10–4	12–4	14–4
Misc. *(a)*-Misc. *(b)*[a]	10–8	12–8	14–8

[a] Miscellaneous combinations include all combinations not specifically provided for. Misc. *(a)*-Misc. *(a)* refers to combinations of the same elements. Misc.*(a)*-Misc.*(b)* refers to combinations of different elements.

MODULANT M-6 Ring schedule

No. elements in ring	*Code*	*No. carbons in ring*	*Code*
1		1	
2		2	
3	12–3	3	13–3
4	12–4	4	13–4
5	12–5	5	13–5
6	12–6	6	13–6
7	12–7	7	13–7
8	12–8	8	13–8
9	12–9	9	13–9
10	12-A	10	13-A
11	12-B	11	13-B
12	12-C	12	13-C
13	12-D	13	13-D
14	12-E	14	13-E
15,+	12-F	15,+	13-F

MODULANT M-6 *Ring schedule*



MODULANT M-7 Ring system schedule

No.	*Rings*	N *rings*	O *rings*	S *rings*	*Benzene rings*	*Miscellaneous hetero rings*
1	7–1	8–1	9–1	10–1	11–1	12–1
2	7–2	8–2	9–2	10–2	11–2	12–2
3	7–3	8–3	9–3	10–3	11–3	12–3
4	7–4	8–4	9–4	10–4	11–4	12–4
5	7–5	8–5	9–5	10–5	11–5	12–5

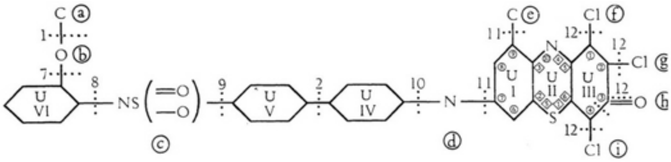
MODULANT M-8 Compound schedule

No.	*Ring systems*	*Saturated rings*	*Unsatd. rings*	*7 and 7 + rings*	N *rings*	O *rings*	S *rings*	*Benzene rings*
1	4–1	5–1	6–1	7–1	8–1	9–1	10–1	11–1
2	4–2	5–2	6–2	7–2	8–2	9–2	10–2	11–2
3	4–3	5–3	6–3	7–3	8–3	9–3	10–3	11–3
4	4–4	5–4	6–4	7–4	8–4	9–4	10–4	11–4
5	4–5	5–5	6–5	7–5	8–5	9–5	10–5	11–5
etc.								

MODULANTS M-9, M-A, M-B Metals schedule

Symbol	*Metal*	*M-9*	*M-A*	*M-B*			*M-B*
Ca	Calcium	8–1		5–1		Group IA	4–1
Cr	Chromium		4–1	6–1	10–1	Group IIA	5–1
Co	Cobalt		6–1	6–1	12–1	Heavy metal	6–1
Cu	Copper	13–1		6–1	13–1	Group IIIB	7–1
Au	Gold	13–1		6–1	13–1	Group IVB	8–1
Fe	Iron	10–1		6–1	12–1	Group VA	9–1
Pb	Lead		11–1	6–1	15–1	Group VIB	10–1
Li	Lithium	11–1		4–1		Group VIIB	11–1
etc.							

APPENDIX B



Symbol	*Meaning*
1,2,3 . . .	Interfix numbers
①②③ . . .	Ring system numbers
◇1,◇2,◇3 . . .	Ring position numbers
I,II,III . . .	Ring identification numbers
ⓐ, ⓑ, ⓒ . . .	Chain unit identification numbers
U	Unsaturated ring
. . . . .	Chain unit demarcation

FIGURE B-1. Preparatory skeletal formula.

THIAZINE CODING SHEET - SPECIES

U.S.P. 1,996,867 (6)

	Card No.	Posn R-R	Interfix 1-6	Interfix 7-12	Fused Face	Sp.		Card No.	Posn R-R	Interfix 1-6	Interfix 7-12	Fused Face	Sp.
S-1	1			11	11-1		S-1						
S-1	1			12	11-1		S-1						
S-1							S-1						
S-2	14						S-2						
S-1	1	4-8	2	10			S-1	1			7,8		
S-1							S-1						
S-2							S-2						
S-1	1	4-8	2	9			S-1						
S-2							S-2						

C.U.	Posn on R 11-	Posn on RS 13-	Relationship O 14-	Relationship M 15-	Relationship P 8-	S-3 Interfix 1-6	S-3 Interfix 7-12	S-4 Interfix 7-12
1.1			14-1			1		
80			14-1			1	7	7
106			14-1				8,9	8,9
64		13-7		15-1			10,11	10,11
1.1		13-9		15-1			11	11
2		13-1	14-1	15-1	8-1		12	12
2		13-2	14-1	15-1			12	12
22		13-3	14-1	15-1			12	12
2		13-4	14-1	15-1	8-1		12	12

Metals

M-9
M-A
M-B

M-9
M-A
M-B

M-9
M-A
M-B

M-9
M-A
M-B

M-8 No. of Ring Collections 4- 4
No. of Saturated Rings 5- 0
No. of Unsaturated Rings 6- 6
No. of 7, 7+ Rings 7- 0
No. of N Rings 8- 1
No. of O Rings 9- 0
No. of S Rings 10- 1
No. of Benzene Rings 11- 5
Acid Salt 13- 0

S-5

S-6 Patent No.

R&D - Thiazine Art - 3/25/58

FIGURE B-2

Symbol identification nos.	*VS_3 Codes*	*Interfix*							
I	M-5 7-9								
	M-6 6-1 11-1 12-6 13-6								
	S-1							11	
II	M-5 4-1 6-1 7-5 13-4								
	M-6 5-1 6-2 11-9 12-6 13-4								
	S-1								
III	M-5 7-9								
	M-6 6-1 11-1 12-6 13-6								
	S-1								12
	M-7 6-A 7-3 8-1 10-1 11-2								
	S-2								
IV	M-5 7-9								
	M-6 6-1 12-6 13-6								
	S-1 4-8		2				10		
	S-2								
V	M-5 7-9								
	M-6 6-1 12-6 13-6								
	S-1 4-8		2			9			
	S-2								
VI	M-5 7-9								
	M-6 6-1 12-6 13-6								
	S-1			7	8				
	S-2								
a	M-C 3-1 5-1 6-1 7-A 9-1 14-1								
	S-3	1							
b	M-3 4-1 5-3 6-1 10-4 14-1								
	S-3	1		7					
	S-4			7					
c	M-1 4-1 5-1 6-4 7-8 8-5 9-1 10-8 14-1								
	S-3				8	9			
	S-4				8	9			
d	M-3 4-1 5-5 6-2 13-7 15-1								
	S-3						10	11	
	S-4						10	11	
e	M-C 3-1 5-1 6-1 7-A 9-1 13-9 15-1								
	S-3							11	
	S-4							11	
f	M-3 4-D 5-A 6-1 8-1 13-1 14-1 15-1								
	S-3								12
	S-4								12
g	M-3 4-D 5-A 6-1 13-2 14-1 15-1								
	S-3								12
	S-4								12
h	M-2 4-2 5-2 6-3 7-2 13-3 14-1 15-1								
	S-3								12
	S-4								12
i	M-3 4-D 5-A 6-1 8-1 13-4 14-1 15-1								
	S-3								12
	S-4								12
	S-5								
	S-6 Patent No. 1,996,867								

FIGURE B-3

Ring I	M-5	7-9	unsaturated homocyclic ring
	M-6	6-1	code number index (benzene)
		11-1	fused-face pattern A
		12-6	6-membered ring
		13-6	6 carbons
Ring II	M-5	4-1	1 nitrogen
		6-1	1 sulfur
		7-5	unsaturated heterocyclic ring
		13-4	nitrogen and sulfur para to each other
	M-6	{5-1	
		{6-2	code index number (thiazine)
		11-9	fused-face pattern D
		12-6	6-membered ring
		13-4	4 carbons
Ring III		Same as Ring I	
R.S.	M-7	6-A	code number index (phenothiazine)
		7-3	3 rings in system
		8-1	1 nitrogen ring
		10-1	1 sulfur ring
		11-2	2 benzene rings

FIG. B-4. Code meanings for phenothiazine

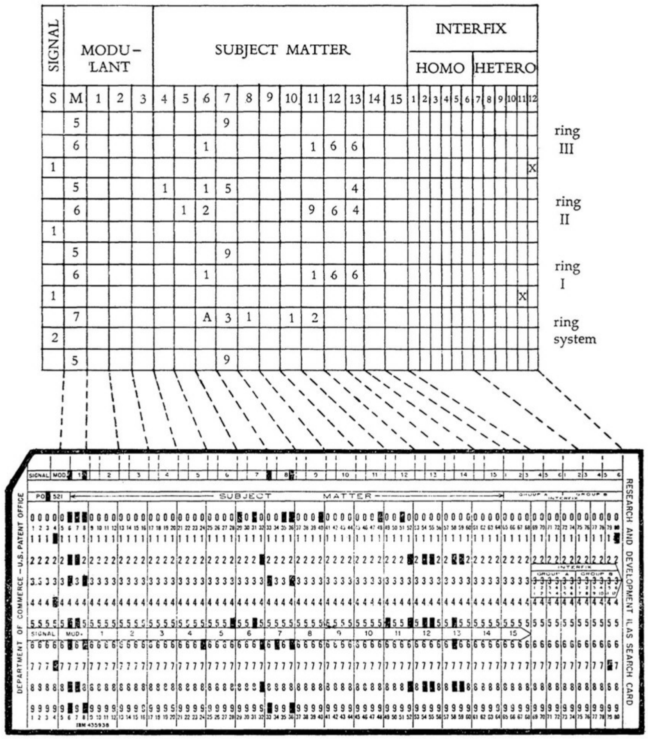


FIGURE B-5. Outline of the punched card used in VS_3. The twelve codes reading from row 9 up correspond respectively to the first twelve codes shown in Fig. B-3

The Haystaq System: Past, Present, and Future

HERBERT R.KOLLER, ETHEL MARDEN, and HAROLD PFEFFER

I. INTRODUCTION

A. BACKGROUND MATERIAL ON PATENT SEARCHING

Literature searching has been distinguished from information retrieval by some authors (1, 5), and within the field of literature searching the peculiar characteristics of patent searching have also been set forth (2–4). The literature contains discussions of various philosophies of searching systems, each of which is derived from a different appraisal of and theory as to users' needs (1, 6, 7). To round out the discussion in the present paper, a few characteristics of patent searching and some of its practical implications are given.

Patent searching is that type of literature searching which is performed by patent examiners when (1) determining the novelty of a concept claimed in an application for a patent; or (2) (if novel) finding the nearest similar related concepts previously known and published. The Patent Office library includes within its files over 2.8×10^6 domestic patents, twice this number of foreign patents, and thousands of serial publications and books. This library literally deals with every field of technology, running the gamut from pins and needles to printing presses, and from antibiotics to submarines. About 2000 searches are made each day in the normal operation of the Office, and each search includes from one to upwards of twenty distinct questions. It is estimated that about 20 per cent of the searches made are in the chemical field, about the same fraction relate to the electrical and electronic arts, and the remaining 60 per cent deal with mechanical and miscellaneous fields. The disclosures of patents in the chemical arts vary in the amount of information they contain from a single, very specific "recipe" to a very generic disclosure of a series of related processes, each illustrated by a large number of examples. The compounds in

HERBERT R.KOLLER and HAROLD PFEFFER Office of Research and Development, U.S. Patent Office, Washington, D.C.

ETHEL MARDEN National Bureau of Standards, Washington, D.C.

volved in a disclosure may be described specifically or generically, and it is common for a genus to be set forth in the so-called Markush type of structural formula. (See Section II.A, "Factors Considered.")

Haystaq is one of many possible types of mechanized searching systems which could have been designed for Patent Searching. Obviously some systems are better equipped than others to satisfy the needs of patent searchers. Some suggested criteria by which systems may be evaluated will be found in Appendix A.

Patent searching is not at all restricted to patents, as the name might suggest. The total field of search does indeed include patents of all countries, but it equally embraces periodicals, textbooks, catalogues, abstract services, and every other form of publication. Many persons other than patent examiners make the same type of literature search; for example, research workers embarking on new investigations, patent attorneys, and lawyers in general.

This type of search is characterized by questions variable in scope from the most generic to the most specific viewpoints. In addition, general combinations as well as subcombinations, equivalence between concepts, negative concepts, and certain syntactical-logical artifices (e.g., Markush formats) are all of importance.

B. HAYSTAQ: SOME GENERAL CONSIDERATIONS

The characteristics of the Haystaq system have been described elsewhere (4, 8); therefore, this paper is principally concerned with new developments and additions to the system. In general, Haystaq includes four parts: (1) a data preparation routine for the library making up the disclosure file of information to be searched (see Appendix D); (2) a data preparation routine for the question, which is set up in the form of a model answer; (3) the search routine; and (4) the checkout routine, which evaluates apparent answers to questions and provides the output of the system. The greatest part of the effort so far has been expended on the search routine since *(a)* the objectives of the two data preparation sections can be achieved manually so long as the system is not yet in a large scale operational phase and *(b)* the checkout routine in extenso is not essential to the searching routine.

Haystaq simulates the manual type of search performed by patent examiners in that it effects a serial scanning of each document in the file. While at present its routines are being coded for the NBS SEAC, full development of the system will be predicated partly on features available only in a more advanced machine. Some of the desirable characteristics of such a machine will be discussed below.

In order to achieve a reasonable working model, the research has been focused on the field of chemistry, although it is quite apparent that minor modifications will permit the inclusion of other fields of knowledge in the system. The file is organized in the most convenient arrangement for the human searcher. Thus, within each document the largest segment of disclosure treated as a unit is a complete *process,* including all the steps disclosed. The next largest segment of disclosure treated is a *composition* or admixture. Each composition is subdivided into groups of codes representing the individual *ingredients* (or "items") which it contains. The individual codes are known as *descriptors.* Since disclosures may vary from simple to very complex statements, any or all of the above levels of organization may be involved.

The system must provide the utmost flexibility with regard to the manner of formulating questions. This need stems from several sources: (1) A very-large file requires a high degree of discrimination to provide exactly (and only) those disclosures which are desired. (2) The same piece of disclosed information can answer a large number of questions, each reflecting a different interest. (3) The ingenuity constantly exercised in phrasing patent claims and the constantly shifting focus of interest in the industrial community result in the continuous generation of new ways of expressing essentially the same or related ideas (2, 3). Examples of types of searches which must be provided for are given in Appendix B.

II. CURRENT WORK

A. FACTORS CONSIDERED

Among the aims of the Haystaq system, as stated in a previous report (8), is the construction of a prototype system. This would permit additional studies as to the requirements and design of a machine system, and provide a basis for constructing improved search systems. Any results obtained from the study of the prototype are a function of the sufficiency of the system, and the sufficiency of the system itself depends upon how well it fulfills the users' needs.

Several problems, set aside for further study at the time of the initial effort, have now been tackled. These are (1) the Markush problem, (2) the creation of an effective classification scheme for the preparation of schedules of items, and (3) the necessity of giving the user greater flexibility in posing questions to the system.

1. "Markush group" is an art term used in the Patent Office for designating a synthetic genus whose scope is determined by a listing of its members. The first system devised was capable of handling this situation where the genus was defined by listing individual compounds, e.g., the class consisting of phenol, mono-chloro-phenol, mono-amino-phenol, and mono-ethyl-phenol. How

ever, another method exists for defining such a group by means of a structural formula showing the portion of the molecule common to all the embodiments, to which is attached a variable member, which is then defined. The same example in the structural format appears in Fig. 1. Where the number of embodiments is small, it becomes a simple matter to decompose such generic formulas into their component embodiments and encode them individually. But it is not uncommon to find such formulas embracing hundreds, indeed thousands, of specific embodiments. One example which was discovered contained over 64,000,000,000 embodiments. It is completely unrealistic to think in terms of decomposing such compact forms into specific embodiments for encoding from the point of view of the number of man-hours involved, the amount of storage required and the time to be consumed by a computer in scanning this mass of data. On the other hand, there are obvious and distinct advantages in being able to compress so much data into an extremely compact form. The problem is serious because the great majority of chemical patents as well as much of the non-patent literature have utilized this device for more than twenty years. The system was therefore designed to process questions in this form.

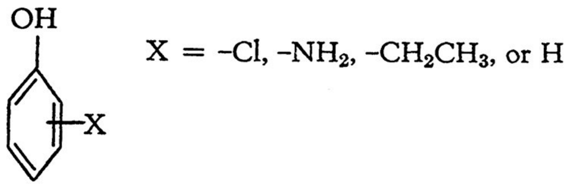


FIGURE 1.

2. Compounds are sometimes defined by generic descriptors alone, sometimes by pure structure configurations alone, and very often by a hybrid of both. Consider the example of the latter situation which is illustrated in Fig. 2.

The terms “halogen,” “alkenyl,” and “alkyl” in a topological network of elements become quite embarrassing in an attempt at tracing an element-by-element path from one portion of the molecule to another. If this example is considered as a question, it is evident that the system should permit recognition of chlorine, bromine, iodine, or fluorine as species of “halogen”; and “propyl,” “isopropyl,” “butyl,” or “isobutyl” as species of “alkyl (3–4 carbons).” In

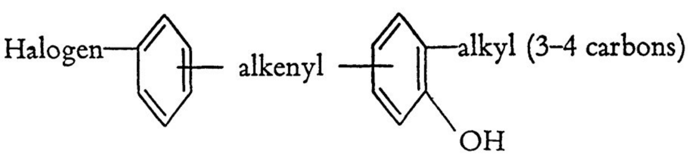


FIGURE 2.

other words, the compound of Fig. 3 should be recognized as an answer. Previous attempts at creating a classification based upon a hierarchical arrangement of generic and specific terms have fallen by the wayside. The large number (practically limitless) of generic and subgeneric terms which can be generated presents such a confused picture of overlapping relationships as to preclude the construction of one comprehensive hierarchy. The usefulness of any comprehensive system will be measured to a large extent by its ability to give full effect to the genus-species relationship.

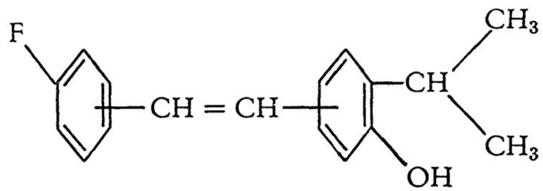


FIGURE 3.

3. The patent examiners, who are among the potential users of the system, are not limited in their searches to mere determinations of the novelty of a compound. Because of legal peculiarities in the patent system, compounds which are similar within certain prescribed limitations are acceptable as answers. Thus, if a question involves the compound of Fig. 4, an acceptable answer may very well be found in any of the structures shown in Fig. 5.

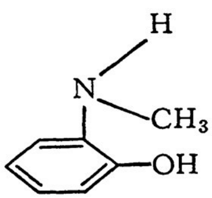


FIGURE 4.

Thus, *5(a)* and *5(b)* are illustrative of positional isomers of Fig. 4 in which all the functional groups are alike, but their relative positions are different.

Figure *5(c)* is illustrative of a situation where all the requirements of the question are met, but the compound has something in addition. If the user required an exact match of the question structure *5(c)* would not be acceptable, nor would *5(a)* or *5(b)*. However, if he specified the question as a fragment of any larger configuration, *5(c)* would be a valid answer.

Figure *5(d)* shows an example of a compound which is a higher homolog of the question, the only difference being in the length of the chain of carbons attached to the nitrogen.

While none of these has the same effect as a direct anticipation, nevertheless they are all valid answers unless they can be rebutted by an applicant.

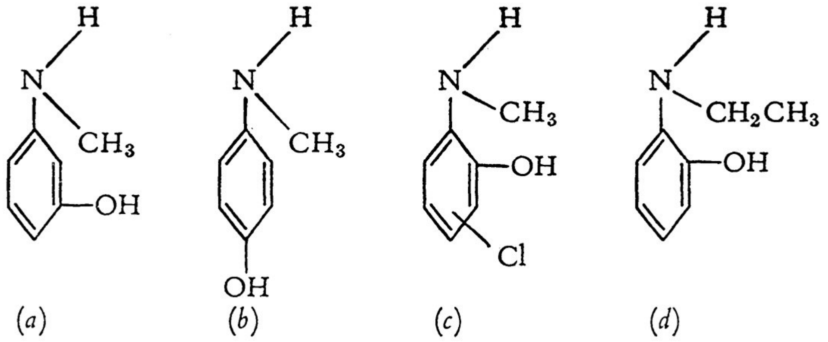


FIGURE 5.

In order that the user be given maximum flexibility in stating his question, the system must permit him the option of indicating, for each point of the molecule in question, whether he will or will not accept positional isomers, whether he will or will not accept homologs and whether he will or will not accept as answers compounds of which the question represents a fragment.

B. THE SYSTEM

In the original search program of Haystaq, chemical compounds were described by a limited schedule of coordinate chemical descriptors. Sufficient definition was thus given to a search request for a specific compound to eliminate a large number of disclosures. This procedure was based upon the anticipated use of a topological element-by-element search, such as the routine written by L.C.Ray (9), whenever required, which would employ the data resulting from the previous operation to give a conclusive answer.

The present system represents an improvement in the search for chemical compounds in the light of the problems outlined above. It is based upon a judiciously selected, comparatively small basic vocabulary of functional groups. These, so far as possible, have been chosen to coincide with the conventional groups recognizable by chemists. In some instances terms have been synthesized, either for convenience in handling or to add additional flexibility to the system.

Inasmuch as recognition of topological relationships among functional groups is one of the desiderata of the system, a close study of two available systems developed by others was made: the Norton-Opler system (10), which deals with relatively large functional groups, and the Ray system (9), which deals with elements. The functional groups chosen for use in Haystaq were selected in an attempt to overcome the relative rigidity of the Norton-Opler system and the relative slowness of the Ray system.

It was not found practicable to use stored look-up tables referring to class terms because of the infinite variety of generic terms and the inability to predict or state all the species which may fall under any given genus. It was therefore decided to make all data, both disclosure and question, self definitive. That is, each generic expression utilized in describing a compound is defined in terms of its specific meaning in that particular compound.

The coded chemical descriptors, all having the initial digit 3, are divided into four sections: 3A, 3B, 3C, and 3D.

The 3A section is devoted to a class of terms which are combinations of two or more units of basic functional groups, having constant definitions and recognizable as such by chemists; e.g., anilino, naphthyl, or carboxyl.

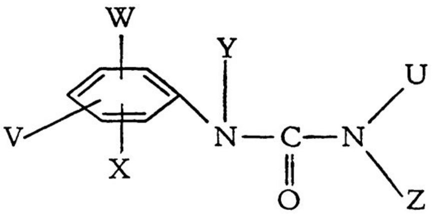
The 3B section is limited to generic expressions, i.e., those capable of being represented by more than one specific embodiment. These are defined, where necessary, by reference to the specific groups in 3C which are pertinent to the particular compound under consideration. Exemplary terms to be found in this group are acid, ester, and halogen.

The 3C section contains a list of all the basic functional groups which make up the particular compound and shows the topological relationships among them (without, however, indicating relative positions of attachment).

The 3D section (not yet written) is an element-by-element topological definition of each compound.

The example of Fig. 6, which is a complex structure having some 1600 distinct embodiments, illustrates a typical structural formula in the literature.

Note that in each case, where H is shown as one of the variants in a group, it is in effect stated that one of the variations is the absence of a substituting group. H is therefore substituted by the expression “N.S.” (no substituent).



U=an alkyl group having one to two carbons; V−H, a halogen, an alkyl group having one to three carbons or an alkoxy group in which the alkyl has one to three carbons; W=−H, or halogen; X=−H, or −Cl; Y=−H, or an alkyl group having one to four carbons; Z=−H, or an alkyl group having one to four carbons.

FIGURE 6.

The structure is rewritten in Fig. 7 in composite form, showing the separation of functional groups, and is arbitrarily numbered for identification of the groups for topological tracing. These numbers are called "designation numbers." Fig. 7 also illustrates the coding scheme.

This coding does not exhaust all possible terms for 3A and 3B; it is only intended to be illustrative. Research is continuing on terminology. In the 3A terms, the number following the description indicates the number of occurrences. Note that all 3B terms (generic) are defined by reference to the groups found in 3C. For example, "halogen-71" in 3B is defined in 3C as a chloro group. The ether of 3B is defined by the 3C terms phenyl, oxy, and alkyl, and is therefore an aryl-alkyl ether.

The first group in 3C identifies piece number 10 as a phenyl group. The first parenthetical expression describes the number of other groups connected to this piece and is expressed as a range. In this case the number may vary, depending upon the particular combination of variable groups which may be present at any one time. Since there are attached one fixed group and three variable groups, each of the latter having as one of its variants a "no substituent" group, the range of connections may vary from one to four.

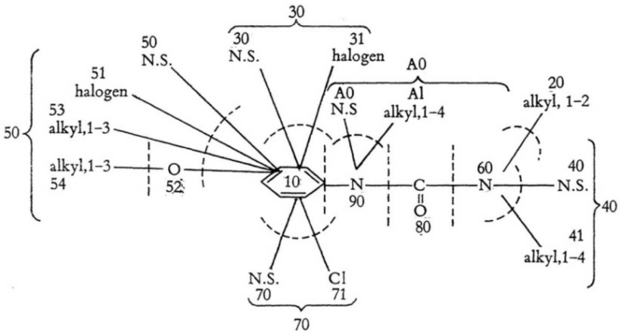
The second parenthetical expression, which has been left blank, is used only for questions and indicates to the computer whether the exact number of (1) or at least as many (0) connections must be matched.

The remainder of the line indicates that piece 10 (phenyl) is connected to each of the pieces numbered 30, 50, 70, and 90 by a single bond (S).

In piece number 30, M indicates the beginning of a variable group. When one of the variants is "no substituent" (N.S.), that information is stored in the M word.

C. SPECIAL CONSIDERATIONS

(a) Ordering. In all the data, in both question and disclosure, the descriptors in 3A and 3B are presented in an ascending series, according to the numerical values of the codes representing the substantive information. The program thus permits the computer to decide (in searching for a particular term) at the earliest possible moment that there is no available answer. For example, if the code for a question term were 127 and the first three descriptors in the disclosure list were 33, 105, and 148, comparison with the first would indicate that it was too small. This would result in calling for the next word, with the same consequence. However, on comparison with the third word, which is too large, the computer would conclude that there is no answer for the question.



3A

anilino—1	urea—1

3B

X-halogen—31	amide—60, 80
X-halogen—51	ether—10, 52, 54
halogen—71	ring—homocyclic, carbocyclic, aromatic, six members—10
amide—80, 90	

3C

10—phenyl (1–4) ( )	—30S, 50S, 70S, 90S
20—alkyl, 1–2 (1–1) ( )	—60S
30—M – N.S.	—10S
X-31—halogen (1–1) ( )	—10S
40—M – N.S.	—60S
41—alkyl, 1–4 (1–1) ( )	—60S
50—M – N.S.	—10S
X-51—halogen (1–1) ( )	—10S
52—oxy (2–2) ( )	—10S, 54S
53—alkyl, 1–3 (1–1) ( )	—10S
54—alkyl, 1–3 (1–1) ( )	—52S
60—amino (2–3) ( )	—20S, 40S, 50S
70—M – N.S.	—10S
71—chloro (1–1) ( )	—10S
80—carbonyl (2–2) ( )	—60S, 90S
90—amino (2–3) ( )	—10S, 80S, A0S
A0—M – N.S.	—90S
A1—alkyl, 1–4 (1–1) ( )	—90S

FIGURE 7

(b) The Question. In the search of Markush type disclosures with Markush type questions in 3C, the computer synthesizes the various combinations of specific embodiments in both question and disclosure. Since this may result in a large number of combinations, some means must be found to relieve the computer of the burden of searching all possibilities. One of the devices employed is the ordering of words in a variable group in the same manner as described above. However, this in itself is not sufficient. It is apparent that for any particular disclosure being searched, not all the variables in the question can be expected to be answered. If fruitless questions can be identified sufficiently early, time can be saved by instructing the computer to disregard them. Consequently the 3A and 3B sections of the question have each been split into two groups, the one containing those terms pertaining to the fixed part of the molecule ($3A_N$ and $3B_N$), and the other containing those pertaining to the variable portion of the molecule ($3A_M$ and $3B_M$). When a $3A_N$ or $3B_N$ term is not matched, the search of that disclosure is immediately terminated. However, failure to find a $3A_M$ or $3B_M$ term results in marking the question 3C pieces referred to, so that they are omitted from consideration when the topological search is made among the 3C terms. The question is re-marked in this manner for each disclosure considered. No terms which are partly fixed and partly variable can be used as 3A or 3B questions. They must all be either one or the other.

If the example of Fig. 7 were to represent a question, the arrangement would be as follows:

$3A_N$	anilino—1
	urea—1
$3A_M$	– – – – –
$3B_N$	amide—80, 90
	amide—60, 80
	ring—homocyclic, carbocyclic, aromatic, six members—10
$3B_M$	halogen—31
	halogen—51
	halogen—71

It is noted that the term “ether” does not appear, since in this example it is in part fixed and in part variable.

(c) Generic expressions in 3C. While generic expressions are ordinarily confined to the 3B section, situations arise, as illustrated in the example of Fig. 7, where a portion of the molecule of a structural formula is described in class terms rather than by some specific embodiment. In this case the 3B terms are repeated in the 3C section, treating them as though they were pieces of basic

vocabulary. Such pieces are marked with an X both in 3B and 3C. In the topological search in 3C, on recognition of such a marked term, the computer is instructed to accept a similar term or any species embraced therein. This is accomplished by a look-up procedure in reverse from the 3C to the 3B section. Thus (referring to the example again as a question), the terms “31-halogen” and “51-halogen” are answered by the terms halogen, chloro, bromo, iodo, or fluoro.

(d) Rings and alkyl groups. The term “alkyl” is of such frequent occurrence that it loses its discriminatory power as a generic expression. This term frequently serves as an additional means of compressing information, as for example “an alkyl group having 1–3 carbon atoms.” The “alkyl” word is therefore given special treatment and has a small subroutine devoted to handling it. The information contained in this word is in fixed fields. One field carries the designation “alkyl.” Another field has information as to the number of carbons involved and is expressed as a range with an upper and lower limit. A third field carries the identification of specific groups. The fourth field is reserved for the question only, and is used to indicate whether the search is for any alkyl group within a specified range or for any homolog at least as large as a specified minimum.

Unsaturated carbon chains are defined in terms of alkyl groups joined through double or triple bonds.

The words which describe rings represent a package of information. Each ring is described in terms of whether it is homocyclic or heterocyclic. If homocyclic, it may be carbocyclic, nitrocyclic, etc. If carbocyclic, it may be alicyclic or aromatic. Heterocyclic rings are described in terms of the kinds and frequency of occurrence of the hetero elements. All rings are further described in terms of total number of elements (i.e., ring size) and double bonds. A special subroutine permits asking for rings in terms of any one or more of the terms described above.

D. THE SEARCH

In general, each of the four sections described may be used as a primary basis of search. However, 3A, 3B, and 3C may each operate as a screen for any one of the subsequent sections. One feature of the 3C section is the creation of an Equivalence Table as part of the output. This table identifies each group of a disclosure molecule which represents an answer for the equivalent group in the question. For example, if Fig. 8 represents a disclosure molecule and Fig. 9 represents a question molecule, the Equivalence Table produced would be as shown in Fig. 10.

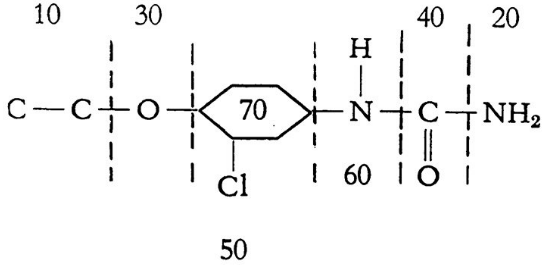


FIGURE 8.

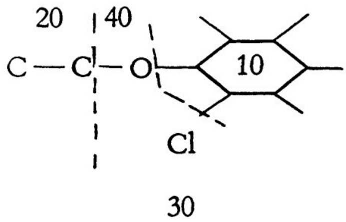


FIGURE 9.

Q		D	
10—phenyl	---------	70—phenyl	---------
30—chloro	---------	50—chloro	---------
40—oxy	---------	30—oxy	---------
20—alkyl, 2–2	--------	10—alkyl, 2–2	--------

FIGURE 10.

Now if the user specifies that no answer is acceptable unless it shows the exact relationship of the substituent groups on the phenyl, the supplemental question for section 3D shown in Fig. 11 provides exact definition. Bv reference to the Equivalence

O — C — C — Cl
40 10 10 30

FIGURE 11.

Table the final element-by-element search spotlights those groups of the disclosure equivalent to question groups 40, 10, and 30, namely disclosure groups 30, 70, and 50. This, of course, has the effect of pointing the 3D search at the critical portion of the disclosure immediately and avoids the needless tracing along innumerable false trails. Since the Equivalence Table is a necessary part of the 3C search section, it costs nothing additional in the way of programming; on the other hand, it proves to be a bonus for the 3D operation.

The search proceeds in serial fashion through each of the sections, the $3A_N$ and $3B_N$ sections acting as screens for the 3C search, and the 3C acting as a screen for the 3D search, when either of the latter is required.

The first step in the program is a check to see whether a 3C search is required, and, if so, whether there is a sufficient number of words in the disclosure to satisfy the minimum requirements of the question. Detection of an insufficient number results in the reading in of new disclosure information. A check is then made as to whether there are any question requirements for section $3A_N$. If there are none the search proceeds to the next section. If a $3A_N$ search is required each question word is matched in turn against the entire ordered list of 3A disclosure terms. The prerequisite for bringing up a new question word is the matching of the previous question word. A failure at this level results in discarding the disclosure and trying a new one.

The next step (assuming satisfaction of the $3A_N$ requirements) is to see whether the question involves a $3A_M$ section. If it does not, examination is made of the question requirements of the next section. If, at the satisfactory completion of any section, it is found that *all* requirements of the *total* question have been satisfied, the document identification is printed out. If a $3A_M$ search is required, the individual question words are matched against the ordered 3A disclosure list. However, a failure at this level does not have the same effect as in the previous section, since these terms represent variable portions of the question molecule. Instead, the question 3C members, identified through their designation numbers, are marked so that they will be omitted in the topological tracing of the 3C section.

The $3B_N$ section is examined after the 3A section of the Question has been satisfied. Each question word is matched in turn against an ordered list of disclosure 3B terms, in a manner similar to the operation in the $3A_N$ section. When a pair of terms is matched, the question is examined to see whether the generic expression itself (the 3B term) provides a sufficient answer or whether definition of the term is called for. Where no definition is required, the operation proceeds as in $3A_N$, where a new question term is brought up after matching the previous one. Where definition of the generic expression is a requirement, reference is made to the functional groups in the 3C section, which either individually or in combined form represent the specific embodiment involved. Comparison is then made of the definitions of the question and dis closure. Failure to match definitions is equivalent to a failure in the $3A_N$ section, and results in bringing in new data for examination. Note that in this situation the generic term acts as a screen, since it must be matched before any attempt is made at checking definitions. Failure of either the screening operation or matching of definitions has the same effect, namely, discarding the disclosure.

The $3B_M$ section is then checked in the same manner as in $3B_N$, but the consequences of failure are the same as in $3A_M$. Since the $3B_M$ terms represent variable portions of the molecule, inability to match them results in the marking of the indicated groups in the 3C section so that they will be omitted from consideration in the topological (3C) search.

Whenever a 3B term in the Question is found to have a special generic mark, no definition is required. If a topological search is indicated, it means that this term appears in the 3C section in place of one of the basic vocabulary units. Whenever this situation arises, all matching 3B terms of the disclosure are stored in a special section of the computer's memory called the Generic Word Storage. If Fig. 7 represents a Question, the matching terms for "halogen— 31" and "halogen—51" are thus stored. When a term marked "generic" is encountered in the 3C section, it is an indication that a corresponding term is acceptable provided it is the same genus or one of its specific embodiments. Thus, in the example an appropriate answer is "halogen" or "chloro" or "bromo." This is explained in greater detail below.

The 3C section follows very generally the logic of the Ray routine (9) in tracing the topological network of the molecular structures. Assume again that Fig. 7 represents a question. The first word in the 3C section is "10— phenyl." The 3C words of the disclosure are scanned until a matching word is found or the end of the list is reached. Failure to match again results in the reading in of a new disclosure. Assuming that a match is found for "phenyl," the number of connections required for phenyl is checked on a consisting or comprising basis as indicated by the question. "Consisting" means that the range of connections of the disclosure must show at least one point of coincidence with the required range, while "comprising" means that the upper limit of the range of the disclosure connections must be at least as great as the lower limit of the question range. If no match is found on this basis, the functional group found is obviously incorrect and another match for phenyl is sought.

Assuming that both the functional group and the number of connections are matched, the information in the first connectivity field of 10—phenyl is scrutinized. This represents the designation number of one of the groups connected to it and the nature of the bond (in this case 30S). The group thus designated becomes the next question descriptor. It is checked first to see whether it represents a single descriptor or the beginning of a variable group (Markush). In this example, 30 is found to represent a variable group. This descriptor is next checked to see whether there is a no substituent variable. Since there is one, the upper limit of the range of connections is reduced by one in the question "source word" (phenyl), and the range again matched against the range of connections shown in the disclosure source word. Assuming that the matching disclosure phenyl group showed three connections, the range on the question is reduced from "1–4" to "1–3" and therefore is still matched by this disclosure. After acceptance of any term, the redundant connection indications are cancelled in both question and disclosure. However, when an N.S. group is accepted, as in this example, cancellation occurs only in the question. Thus, "30" is cancelled from word 10 and "10" is cancelled from word 30. The next field in 10—phenyl is then examined and word 50 is brought up as the next question. This group also proves to be variable and since it also shows an N.S., the range of connections in phenyl is reduced to "1–2." An attempt to rematch against the disclosure reveals that the range is too low. Therefore, this combination of variables cannot find an answer in this disclosure. The range in the question is therefore restored to "1–3" and another member of the variable group is brought up as the new question. This is "X-51—halogen." Recognition of the generic mark (X) activates a look-up routine to discover whether *any* designation number of the answering phenyl group is to be found among *any* of the words of the Generic Word Storage. If such a word is found in the Generic Word Storage, it is compared on a substantive basis with the question, and if they agree, the word in the Generic Word Storage is marked to show which question word it answered. The 3C word which defines this generic word is then substituted for it and the operation continues. Redundant connection indications are cancelled from both question and disclosure. If no corresponding term is found in the Generic Word Storage, another member of the group is tried until all members are exhausted. Should a member with an "omit" mark be found, it is ignored and the next member examined.

Either of two conditions may obtain at this point: a match may or may not have been found for "X-51—halogen." If a match has been found, an attempt is made to continue on the selected trail by looking for other connections from the halogen. If there remain no unmatched connections on this group, the search backs up to the first group that can be found which has uncancelled designation numbers and attempts to proceed forward from that point. In this case, the question word "X-51—halogen" has no further connections; therefore the backup is to the preceding word, "30—M—N.S." Since that word has no further connections, another step back is made to "10—phenyl" and the next unsatisfied connection, "70S" is selected. The search then proceeds in a forward direction again. As progress is made forward and backward in the question, there is a corresponding movement through the disclosure. Thus, for example, when the search backs up to "10—phenyl" on the question side it also backs up to the corresponding word in the disclosure. When all selected question words which have been matched have all their connection designations cancelled, it is established that an answer has been found.

Now consider the condition that no match has been found for "X-51— halogen." On testing for another member of the 50 group, it is found that all members have been tried without finding a match for any of them. This was therefore a blind trail. It is possible that if another answer is found for the 30 group (which was the first connection attempted from "10—phenyl") an answer may appear for the 50 group. The search therefore backs up to the last place that matched and attempts to find another answer. In the example given, the last accepted connection was for an N.S. member of a variable group; therefore, the range in the parent or source word is restored and the next member of that group is tried. (If this word is anything other than an N.S., an attempt is made to find another answer among the remaining connections in the disclosure word of the last match.) If none of the other members of the variable group can be matched, it is an indication that the apparent answer for the source word was incorrect. Therefore, another attempt is made to find an answer for "10—phenyl." Since this is the first word, the search proceeds down the list of disclosure words, continuing from where the last apparent answer was found. If another phenyl word occurs, the search again attempts to trace through the structural network from this point. If no other match for phenyl is made, a new disclosure is brought in. In general, when backing up to look for new trails, the attempt is made to find other answers for connections from a source word which already has been matched. When all such attempts prove fruitless, a different match for the source word itself is sought.

E. STATISTICAL

Use of the present word formats requires an average of two computer words for each descriptor. A SEAC word contains forty-four binary digits plus a sign digit. The example illustrated in Fig. 7, with its 1600 specific embodiments, was completely coded with fifty-two SEAC words.

Appendix C describes several test searches made by use of a manual simulation of the machine search.

F. STATUS OF THE SYSTEM

The generalized Haystaq program, which defines the search parameters, has been coded and debugged on SEAC (8, 11). It provides for searching com pounds, admixtures, and processes. Amplification of various areas of the program is now in progress. The program for searching chemical compounds in terms of functional groups, which has been described above, appears to have utility per se as a means of searching for individual compounds. By a slight modification, functions or uses may also be included. It is expected to make it an integral part of the generalized search program. At the time of writing, a flow chart for this part of the program has been completed and the machine instruction code is being written. It is hoped that significant results will be available for a report at the time of the Conference. Additional areas for future amplification include the element-by-element topological search (9) and the process search (3, 4). Tests of the searching system must be made to evaluate not only the logic of the system, but also the sufficiency of the various techniques provided in terms of how well they meet the users' needs.

III. LOOKING TO THE FUTURE—A LONG RANGE VIEW

The second paragraph of the introduction gives a clue to the magnitude of the search problem in the U.S. Patent Office. A mechanized searching system which will meet all the needs of the Patent Office must make provision for satisfying the following criteria:

1. The ability to conduct a large number of searches every day, with no long delay interposed between question input and answer output.
2. An output convenient for direct application by the user of the search system. This implies *(a)* a convenient physical form of output and *(b)* no “false drops.” (False drops are wrong answers which result from ambiguities in the system.)
3. Files which can *in effect* be completely searched for each question. (See the section below entitled “Machines with Learning Ability,” part 1.) For the Patent Office, this has implications for the presumed validity of patents. For other users, the completeness of a search may bear on the question of whether a research project is warranted insofar as it does not represent a duplication of another person's earlier work.
4. Expansibility of the system to permit insertion of all types of information into the library file and to permit relatively easy entry of newly created disclosures. The system must be comprehensive enough to permit any question to be asked, regardless of subject matter.
5. A relatively inexpensive and simple system, both as to inception and operation.
6. Ability to satisfy all the needs of the users. That is, all logical types of questions should be capable of being formulated and the series of answers put

out by the machine should reflect no "noise." Inability to state a question precisely because of limitations in the system results in a broader or narrower question than is really intended. In the broader case, non-pertinent and therefore undesired answers, which constitute "noise," are developed; in the narrower case, some answers may be lost.

A system of the magnitude required which can meet these criteria can result only from a concerted effort over a long period of time.

The argument that a large, complex searching system geared to the needs of large-scale users, such as the Patent Office, will not be the best solution for small-scale users, who might be satisfied with specialized but simpler systems, is analogous to the old argument for large, fast, expensive general-purpose digital computers vs. smaller, specialized machines. Since the Patent Office is concerned with all of technology, the Patent Office problem inherently includes every one's technical literature searching problems. Even the Patent Office does not contemplate *literally* searching all its files for every question. (See the section below entitled "Machines with Learning Ability," part 1.)

Many disciplines, seemingly unrelated to mechanized literature searching, must be developed far beyond their present state in order to provide some of the tools which will be used in formulating "the ultimate system." Several related fields are briefly discussed.

A. PRINTED DOCUMENT READING

One of the most voluminous and demanding tasks in setting up a large searching system will be the preparation of the search file (the library file). An enormous expenditure of manpower will be required to analyze and encode the documents and to transcribe the codes onto a permanent storage medium. Since the effectiveness of the system will depend largely upon the accuracy and completeness with which the file is prepared, the importance of this task cannot be overemphasized. This job will ultimately have to be performed by machines if our long-range objectives are to be achieved.

Much work is now in progress in the design of systems of "character recognition" by machines. While most of the effort now being expended in this field is concentrated on the development of equipment which can recognize individual printed or written alphanumeric characters, progress is advancing to the point at which printed words can be recognized by machines and stored in their memories in an encoded form (12, 27).

B. THE MECHANICAL TRANSLATION OF LANGUAGES

This is also the subject of many investigations (13). It requires little vision to foresee a system incorporating the techniques of both the character recog nition and mechanized language translation arts. A printed document might be the input for a machine which "reads" the printed page, analyzes the words, sentences, paragraphs, etc., for their meaning and translates them into an encoded form of the original document suitable for machine searching. It is hoped that within the next decade such a scheme for preparing a mechanized searching file will be a reality.

In Appendix D there is set forth one prospective means for using machines to accomplish certain portions of the data preparation task.

C. MACHINES WITH "LEARNING" ABILITY

The concept of machines with "learning" ability has been the subject of much discussion in the past several years (14–16).

Three types of procedure which might be classified as involving *quasi*-learning by a searching machine are described:

1. Methods which enable the machine to make what are in effect complete searches without actually searching each disclosure in the file

(a) Use of the machine to break down a large file, into smaller classified files. Some people assume that a large file of documents encoded for machine searching should be physically separated into smaller classified collections in order to save a large amount of the time required for searching the entire file. This assumption should be considered in the light both of the nature of Patent Searching and the inherent limitations in "classification systems" which employ any of the schemes in which the units to be classified are entire documents.

This type of classification system generally involves a scheme, as in most manual classification systems, for physical placement of copies of documents. R.A.Fairthorne refers to such a scheme as "Parking" (17). Any scheme which makes use of document classification involves some sort of arbitrarily established rules of superiority. The ability to select a pigeonhole for a document has a certain nicety about it for the classifier. However, the searcher is frustrated in his attempts to retrieve all documents disclosing idea A because of a rule which states that if idea B is disclosed it takes precedence over all other ideas, including A. As a result, those documents which disclose both A and B are hidden from him. Cross-referencing to the A location is a partial solution, but, as will be shown hereafter, it is a practical impossibility to cross-reference any document sufficiently to provide for the needs of all searchers.

Some general characteristics of disclosures of documents, considered from the point of view of Patent Searching, are as follows:

(1) All subject matter details of a disclosure are equally important because

any one or any combination of them may be the basis for a search. Subject headings are completely inadequate for the patent searcher.

(2) All disclosed relationships among the small units of information are equally important for the same reason. Examples of such relationships are combinations, and all their included subcombinations, disclosures of equivalence, and negatively disclosed ideas.

(3) Each concept disclosed may be described (and therefore should be approachable in searching) from a great many points of view, and each of these points of view has both generic and specific aspects.

In the light of all these considerations, a rigid classification of entire documents does not suit the needs of the patent searcher, since it inhibits his ability to synthesize the desired classification for each search. It should be noted that an enormous amount of labor is required to keep rigid document classification systems current. Havoc results (in apparently sound schemes) when new developments shift the emphasis from one phase of an art to another.

The argument for physically separating a large file into smaller files is based on the existence of broad, mutually exclusive, categories, such as chemistry and the mechanical arts. Should one be forced to examine a total document collection, including such disclosures as clothes washing machines, when his interest lies in a particular synthetic resin? Surprisingly (only to the uninitiated), the washing machine art may very well be a source of disclosures of synthetic resins which are used in the protective coating for the casing of the machine, or in the manufacture of the agitator element, the tub, the gears, or other parts, and these resins may be the very ones sought. It is impossible to estimate accurately the number of such hidden disclosures, but it is certainly a large number. All who have had experience with manual searching systems are familiar with the accidentally discovered disclosure found in an unlikely place in the classification scheme.

One plan which permits the mechanized searching of files which are smaller than the total collection would begin operations with a single, all-inclusive file. Records would be kept of questions searched by the system. As the total file grows (by the incorporation of additional documents into the system) to such a size that searches become burdensomely long, the records of previously asked questions could be analyzed to discover frequently asked questions involving common types of subject matter. It then becomes feasible, as the need arises, to set up a separate file of synthetic resin disclosures. This would be accomplished by using "synthetic resins" as the question, and having the machine copy all documents disclosing such subject matter onto another storage unit. The copied documents remain in the original file and continue to be available for search on the basis of other subject matter which they disclose. All future searches for synthetic resins would be made on the smaller file only. Such searches would be equivalent to a complete search of the total file. Almost every search involves some concept which was never before made the basis of a search. A document answering one search may be completely irrelevant for another, and, conversely, those documents which were completely non-pertinent for all previous searches may be exactly the ones desired in a later inquiry.

(b) Learning by experience. A list of questions which have been asked previously can be stored, together with their answers and an identification of the last document inspected in the previous search. In making a search, the "previous question" file would be scanned first. If the question asked proves to be an old one, the previous results are printed and the search brought up to date by searching all subsequently entered disclosures. Where the question is a new one, the entire disclosure file is searched and the new results added to the stored list. This saves searching the entire storage file for every question. This proposal has utility only until such time as the search of the "previous question" file becomes unduly burdensome.

(c) Association of related documents. Encoded disclosures can contain information which leads from one disclosure to all related disclosures and therefore detailed searching is required of only those documents which have some pertinence to the question asked. Suppose that all documents containing the same information are figuratively arranged as points along a line. Following each code in each disclosure an identification of the adjacent document in the line would appear. Where two search criteria are specified, the search is for a disclosure which is at the same time a member of (1) the line representing all disclosures containing the first code and (2) the line representing all disclosures containing the second code. It is thus the intersection of these two lines that is sought. Since two codes may be desired in a single disclosure only in a particular relationship, it becomes necessary to select intersections of a specified type representing that desired relationship. For example, if the search is for A and B in the same composition, either A or B can be used as the initial basis for search. On locating the first document that discloses A, search can then be made in that document for the additional presence of B in the specified relationship to A. If B is not found, the information in the code for A will identify-the next document that includes A, and that document is searched for B. Should one document be located which fully answers the query, the "next document identifications" in both A and B are compared. The search proceeds to the one farthest removed. In other words, if the next B document is farther removed than the next A document, the B document is examined next and B becomes the primary basis for search, with A as the secondary search criterion. This method of searching involves locating one line which is perti

nent, then examining only members along that line until a proper intersection between the correct lines is located.

2. *Methods which enable the machine to find answers which the human searcher did not preconceive*

In manual searching, when it is desired to find A and B in a particular relationship X (e.g., A is a fly killer and B is a solvent and the relationship is that they are in admixture in an insecticidal composition), the searcher is aware that such a specific disclosure may not be found in one reference. He therefore looks also for disclosures of A in combination with materials that are equivalents of B, as well as for disclosures of B in combination with materials that are equivalents of A. A reference disclosing A and C in X relationship, or one showing D and B in X relationship, is called a "basic reference," or "one showing the general or basic combination." A reference showing the equivalence of C and B in a G relationship, or one showing the equivalence of A and D in an H relationship, is called a "secondary" or "supporting" reference. If the requirement for A and B in an X relationship cannot be found in a disclosure, a combination of a basic reference and a secondary reference must be used.

A program could be written for a machine as follows:

(1) The primary question, "Find A and B in the X relationship," is given to the machine by the user.
(2) The machine generates these additional questions: *(a)* Find A and anything having the function of B (e.g., Y) in the X relationship. *(b)* Find a statement of equivalence between anything (e.g., Y) and B, where equivalence means that they have the same function, and that function is the same as in *(a)*. *(c)* Find B and anything having the function of A (e.g., Z) in the X relationship. *(d)* Find a statement of equivalence between anything (e.g., Z) and A, where equivalence is based upon the function of A, as expressed in *(c)*.
(3) Initially, the machine scans each document to determine whether it answers any of questions *(a)* through *(d)*.
(4) If *(a)* or *(c)* is found, the machine determines whether the primary question in (1) is answered; that is, that Y is actually B or that Z is actually A.
(5) If the primary question is answered, the machine continues to search for answers to the primary question only, dropping questions *(a)* to *(d)*.
(6) If the primary question is not answered in (4), the search continues as in (3).
(7) The machine prints separate lists of the documents found as answers for each question.

(8) If there are no answers to the primary question, the machine correlates answers appearing on the *(a)* list with those on the *(b)* list to find all pairs of documents where Y is the same; a similar correlation is performed between the *(c)* and *(d)* lists to find those pairs of documents where Z is the same.

3. Methods which enable the machine to supply the best answer it can when a set of alternative questions is asked.

When it is desired to make a very specific search, the searcher keeps in mind the possibility that the exact answer he wants is not available. He therefore poses search questions which vary in scope from his specific question up to the broadest question which will satisfy him. Thus, he sets the outside limits, both most specific and most generic, which satisfy his needs, and he may include certain searches of intermediate scope as well. The most specific search could be made on the total document file, and should no answers be forthcoming a broader search could be made. This process would be repeated until the broadest satisfactory search is completed. If answers to any of these questions are found, no further runs are made, since the answers retrieved are the most specific available. Such a system of alternately asking questions of and receiving answers from the machine has been aptly called "conversing with the machine." This would be a relatively time-consuming method of operation.

A more satisfactory approach, which would permit the machine to carry on a "monologue," would operate in the following manner. Let a series of questions, A, A′, A″, A‴, and A″″ (listed in order of increasing desirability), be propounded at the beginning. The order does not necessarily reflect either an order of variation of the size of a combination or an order of variation of genericity.

Searching would begin with question A. If an answer to A is found, a determination is made as to whether it also satisfies questions A′ or A″, etc. Assume that the first reference found only answers A. The identification of the document is recorded and searching continues. Suppose that the next answer found satisfies A, A′, and A″. Then the answer previously found is discarded since a more desirable answer has been obtained, and questions A and A′ are dropped. The search continues on the basis of question A″. On locating the next reference, determination is made as to whether it satisfies only A″ or whether it also satisfies the more desirable A‴ or A″″. If eventually A″″ is located, the only answers given as output reflect A″″ disclosures, but if A″″ is not found, the most nearly desirable answer available is given. By using this approach, in one pass through the stored data answers would be obtained within the defined limits of the search as near the most desirable answer as is available.

D. THE FIELD OF LINGUISTICS

Several projects under way are concerned with structural linguistics (18–20). Analysis of sentence structure is a useful technique for determining the intended meaning of an ambiguous expression. In addition, there are many problems relating to the recognition of a single concept whenever it is expressed in diverse ways.

The development of a large-scale retrieval system will depend largely upon the results of long-term research in concept classification. That word classification has little significance is apparent when it is recognized that the “natural language” of printed documents will seldom, if ever, be found to be the best language for machine use. Languages are, after all, tools for the symbolic expression of ideas, and each language uses its own symbolism (the vocabulary) and grammar (rules for concatenating ambiguous vocabulary symbols to express particular, distinguishable concepts). It would be unexpected to find two such different kinds of entities, as are men and computer-like machines, “thinking” in the same language.

However, a humanly useful output from any machine system must be in a language which is intelligible to the human. In the case of a literature searching system, this might be in the form of a secondary type of output such as a bibliographic listing or an abstract, or it might be in a primary form, such as a printed copy of the document, a microfilm, a cathode ray tube display, a magnetic tape for audio presentation or the original document itself. Empirical bites into the linguistic problems have been made (in the program described in Section II) by the use of internal definitions, “look-up” features, and the possibilities of using unlimited numbers of 3A and 3B terms.

E. SOME MACHINE DESIGN PROSPECTS

In general, Haystaq can be regarded as an example of one general approach to a large-scale, mechanized literature searching system. As previously stated, it simulates to a large extent the search performed manually by a human searcher. No economic feasibility studies have as yet been made since this system is still highly experimental. It has been primarily an exercise in developing methodology in a field in which no guiding generalities have been established.

Large-scale systems which would be completely satisfactory are probably a considerable way from us. It is of interest to note that in one very large machine manufacturing corporation, research projects are considered to be short range if their probable duration is as much as three to five years. There are pressures to effectuate crash programs for solving the information access problems. These must be deferred to, but the restricted long-term prospects for narrow-gauge searching systems must be kept clearly in mind (21). As has been the usual story in technological developments, not only must we walk before we can run, but we must also crawl before we can walk.

A look at probable trends in the foreseeable future and possible trends in the years further ahead follows. For the time being, we shall have to be content with machines that, in general, are improved versions of the presently known general purpose computers. These have been designed to handle either (1) mathematical problems (the so-called scientific computers) or (2) business management problems (the so-called data processing machines). The operations which can be performed are generally the same in either case, but the facilities provided in the former emphasize internal operating ability (since a relatively large amount of computation is performed on relatively small amounts of input data in mathematical problems), whereas the latter place more emphasis on input and output (since business problems generally require a relatively small amount of computation on each of many units of input data). However, since many problems are coming to light which require both large amounts of input and output equipment and fast internal operating speeds, the distinctions between scientific computers and data processing machines are not so important as formerly.

In contrast to these problems which deal with quantitative information, the literature searching problem requires a machine designed primarily for handling qualitative information. Recognition of specified small units of information and the complex webs of relationships among such units must be among the primary functions of this machine. Any computation performed in searching will be purely incidental. What is desired primarily is not arithmetic calculations but flexible matching ability. The “orders” available on a searching machine should be tailored to the job they are to do. Searching programs written for existing machines are handicapped by operating under the guise of mathematical processes, and this procedure involves many red-tape operations.

Present-day computers operate in a serial manner, one piece of data being treated only after another has been processed. Various schemes have been employed in different machines in an attempt to overcome this handicap. These include (1) overlapping the non-productive portions of two successive operations, (2) providing for simultaneous input, output and computations, (3) enabling the machine to carry out fully independent operations simultaneously in different parts of the computer and (4) providing faster operating input, output, and computing components. Until machine designers can provide us with real searching machines (and it is the responsibility of documentation specialists to educate the machine designers on the characteristics desired in such equipment) all these techniques must be used.

A machine system which would facilitate searching (based on the present type of general purpose computers) might include a cascade of machines, employing non-fixed word lengths, having a choice of types of output, ultra-high speed input and internal operation, and certain types of quasi-learning features.

The cascade of machines referred to would employ special searching mechanisms for each of several parts of the search. Each machine would operate at a level of intensity somewhat greater than the previous one. The documents selected by the first machine on the basis of very broad criteria would become the input for the second machine. The output of the second machine (selected on a more intensive basis than the output of the first) would become input for the third, and so forth. For example, disclosures selected on the basis of general chemical criteria would be fed into a machine specialized for making topological structure searches. The output from this stage would become the input for a checkout routine machine. The output from this last machine would be fed into a microfilm reproducing machine to supply the user with the ultimate output of the system. If the criteria for the several specialized machines were properly selected, all machines could be kept operating in a gear train type of synchronization, the speed of operation of each machine being balanced against its own workload.

Ultimately we will have methods of searching literature by means of a fully mechanized system. Machines will be given copies of printed documents and will then proceed to prepare their own encoded library files. Searchers will be able to put their questions to the machine orally in a common, natural language. Much learning ability will be inherent in the system, so that as they become more experienced the machines will search more efficiently. The searching machines will be parallel machines, born and bred for their job. (See Appendix E and paragraph (j) of Appendix A.) At that time, we will have available an operating embodiment of the "automatic library" envisioned by Vannevar Bush (22, 23).

APPENDIX A EVALUATION OF LITERATURE SEARCHING SYSTEMS

Mechanized literature searching systems may be described from many points of view. Descriptive factors may relate to *(a)* characteristics of the original document collection, including its physical attributes as well as the types and scope of included information, *(b)* the methods and tools for classifying, indexing, and encoding the document collection to prepare the search file, *(c)* characteristics of the search file, *(d)* the formulation of search questions, *(e)* searching and retrieval techniques, *(f)* the output of the searching operation, *(g)* the updating of the search files, *(h)* the searching machines, and *(i)* personnel requirements and economic factors.

Comparisons among systems must be based largely on their effectiveness in fulfilling the needs for which they have been designed. The success of a system, from this point of view, will depend upon several criteria. (1) The size of the file and the extent of its heterogeneity balanced against the frequency of use of the system and the complexity of question types will determine the degree of discrimination or resolving power necessary to separate wanted from unwanted disclosures. (2) “Noise,” resulting from undue breadth in the permissible questions and “false drops,” resulting from ambiguities in the system, can be tolerated only to the extent that they do not detract from the utility of the output to the user and to the extent that they do not deter the user from formulating the exact questions to which he must have answers. (3) The amount of detail used in describing concepts disclosed in the document file of a searching system will depend upon the needs of the users. (4) Sufficient flexibility in the logical types of questions permitted must be provided to satisfy users' needs.

It is believed that when a retrieval system is to be designed, the following questions must be considered together with points (1)–(4) above.

(a) Should the main approach to subject matter be on a statistical (i.e., probability) basis (as in all coordinate indexing systems) or on a precise basis? If the system is constructed on a statistical basis, how much redundancy is desirable? For re-entrant systems [see *(h)* below] using serial computers, a combination of statistical factors for “screening” followed by the use of precise factors for the ultimate selection appears to be desirable.

(b) Should the organization of subject matter be based upon discrete units at varying levels of intensity or is a topological network of all disclosure factors preferable? Again, a combination of these approaches appears to permit question flexibility so that a user may ask his question in the form most indicative of his actual interest.

(c) Should a limited number of “subject headings” (telling in broad terms what the document contains) be used, or should all subject matter details be included? Because a system designer cannot predict exactly what questions will be asked, he must provide for a multilateral approach to each document and make all the disclosed information details equally accessible.

(d) Are the *words* used in a disclosure adequate to describe the subject matter, or must the *concepts* implied by the words of the disclosure be made available for retrieval?

(e) Although the primary concern is for a literature searching system, will there also be a demand for such uses as “Information Retrieval” and “Browsing?” Any

literature searching system employing "encoding-in-depth" inherently includes in the search file the kind of data useful in an information retrieval system. A modification only of the type of output would be required to employ the file in this dual role. Browsing is based upon a philosophy which is basically different from searching or information retrieval, but the nature of search files is such that a searching system can be designed to provide facilities for browsing.

(f) How much "noise" and what per cent of "false drops" are tolerable? It is not true that a large-scale searching system must have some degree of each of these disturbing characteristics.

(g) Is it necessary to store the disclosures so that the original document can be reconstructed, or is it adequate to provide clues to the subject matter in the documents being searched? In part, the answer to this question depends upon the form of output provided and the use made of the output.

(h) To what extent should the system be a re-entrant one? This refers to the ability of the machine to retain, manipulate, and compare various pieces of information presented to it at different instants of time. The answer is largely determined by the complexity of questions permitted and is limited by the type of searching machine used. Questions *(a)* and *(b)* are related to this one.

(i) Should the file be static or kinetic? Hand-operated notched-edge card systems exemplify the former type, while any of the systems employing the ESM—101 with IBM cards is an example of the latter. Factors such as the time to make a search, the complexity of questions permitted, and the possible size of the file must be analyzed.

(j) Should the system employ a serial or a parallel approach? No sound theoretical reasons exist which substantiate the passive acceptance by many documentation specialists of the concept that it should take more time to search a large file than it does to search a small one. Only the present-day absence of machine technology for large-scale parallel searching is lacking. Some systems provide for several questions to be asked simultaneously when one serial pass through the file is made (24, 25). The parallel system contemplated here does not involve parallel questions; what is contemplated is a parallel or simultaneous approach to all documents in the file. (See Appendix E.) In such a system, no prolonged interval need exist between the instant when the question is put into the machine and the instant when the output is obtained. Questions could be asked one after another at the same rate of input as data is fed into present-day electronic computers.

(k) What types of logical subject matter systems must be devised to permit the formulation of all desired questions?

(l) What are the costs involved in setting up the system? Initial costs include document accession, system development and programming, data file preparation, machine acquisition or rental, and the employment and training of necessary personnel. Initial outlays should be amortized over a reasonable operating period. A large first expenditure may be less costly in the long run than a smaller first cost for a system requiring extensive maintenance.

(m) What is the cost per search for operating the system? The cost of updating the files, which must be considered as a continuing operation, must be added to the expense of performing a search.

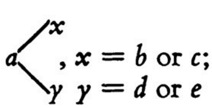
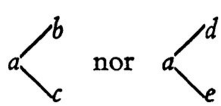
(n) In devising and setting up a full-scale system, how much time will be required before productive operation is possible? A system conceived along superficial lines

will lend itself to earlier exploitation, but will require more effort in the later incorporation of desirable features than if they had been included initially. Careful planning plus the provision for more detail and more logical ability than appear necessary in a first appraisal will generally be worth the investment in time required.

(o) How long does it take to make a search and is that length of time satisfactory? Formulation of the question in machine language, programming the machine, actual running time, and obtaining the output in a usable physical form must all be taken into account.

(p) How difficult is it for a user to frame a question? If the user must master a complex coding system before he can make use of the system, he is less likely to accept it and use it to best advantage. An educational program to acquaint users with the facilities provided by the system is required. A practical method of operation permits the user to state his question in common language, and a technically trained person, who is intimately familiar with the system, translates the question into the system language.

APPENDIX B SOME EXEMPLARY TYPES OF QUESTIONS

1.	*A* or *B* or *C*	where *A, B,* and *C* represent different specific compounds, any one of which will satisfy the question.
2.	*A+B+C*	where *A, B,* and C represent different compounds which must be disclosed in admixture.
3.	*abc*	where *a, b,* and *c* represent different aspects of a single compound.
4.	*(a) abc* or *def*	similar to type 1, but each compound is described in terms of several aspects.
	(b) abc or *abd*	where the two compounds have some aspects in common.
5.	*(a) abd+def*	similar to type 2, but each compound is described in terms of several aspects.
	(b) abc+abd	where the two compounds have some aspects in common.
6.	*ab* (without *c*)	similar to type 3, but one of the stated aspects must be *not* present.
7.	*ab* (no *c*)	similar to type 6, but the disclosure must positively exclude aspect *c*.
8.	*a* (without *b*)+ *b* (without *a*)	similar to type 5, but the two compounds have no common characteristics.
9.	(*A* or *B*)+(*C* or *D*)	where *A, B, C,* and *D* represent different compounds. Note that neither *(A+B)* nor *(C +D)* is requested.
10.	*abc* or *ab*	First choice *(abc)* is a more specific concept than the second choice *(ab)*.
11.		This is an example of a Markush structural formula. Neither  is a valid answer.

Types 12–15 represent processes, where *A, B, C,* etc., represent the various materials present in the different steps of the process.

12. $$(A+B)\rightarrow(C+D)\begin{cases}\nearrow(C)\\ \searrow(D)\end{cases};\ (C+E)\rightarrow(F)$$

13. $$(A+B)\rightarrow(C+D)\begin{cases}\nearrow(C)\rightarrow(E);\\ \searrow(D)\rightarrow(F);\end{cases}\quad(E+F)\rightarrow G$$

14. $$(A)\rightarrow(B)\rightarrow(C+D)\begin{cases}\nearrow(C);\ (C+E)\rightarrow(F+I)\rightarrow(J)\\ \searrow(D);\ (D+G)\rightarrow(H+K)\rightarrow(J)\end{cases}$$

15. $$(A+B)\rightarrow(C+D)\rightarrow(A+E)\begin{cases}\nearrow(E)\\ \searrow(A)\end{cases}\quad[(A)\text{ returns to }(A+B)]$$

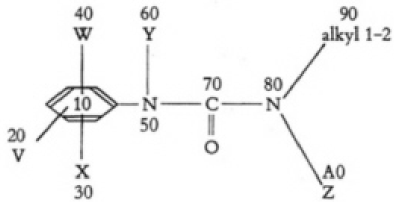
APPENDIX C TEST EXAMPLES OF HAYSTAQ RUN ON SIMULAC

The logical sufficiency of the flow chart for the system described in Section II was checked by manually simulating the activity of the computer on several test problems, each involving Markush structures. During such tests several errors in housekeeping operations were discovered and corrected. Code checking time on the machine is expected to be substantially reduced by this method of pretesting. Simulac is the simulation of an automatic computer on a blackboard.

TEST 1

(Question 1 has 1600 unique embodiments, Disclosure 1 has 48 unique embodiments, and hence the search of this disclosure by this question is the equivalent of 76,800 separate searches.)

Question 1:



V = N.S. (No Substituent) [20], alkyl 1–3 (i.e., C_1 to C_3) [23], halogen [21], or alkyl 1–3 [24] oxy [22]

W = N.S. [40] or halogen [41]

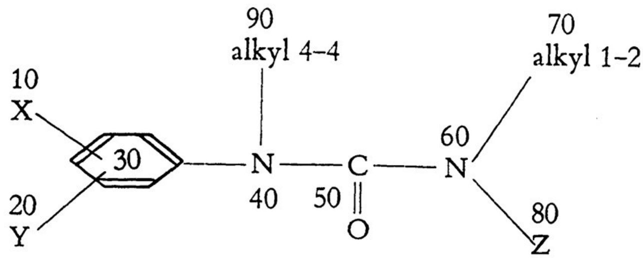
X = N.S. [30] or halogen [31]

Z = N.S. [A0] or alkyl 1–4 [A1]

Y = N.S. [60] or alkyl 1–4 [61]

FIGURE 12.

Disclosure 1:



11 12
X = alkyl 1–3 or NH_2

21 22
Y = fluoro or OH

80 81
Z = N.S. or alkyl 1–2

FIGURE 13.

Simulated print-out, test 1: Question embodiment:

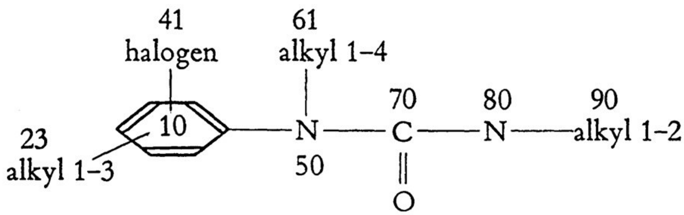


FIGURE 14.

Satisfied by Disclosure embodiment:

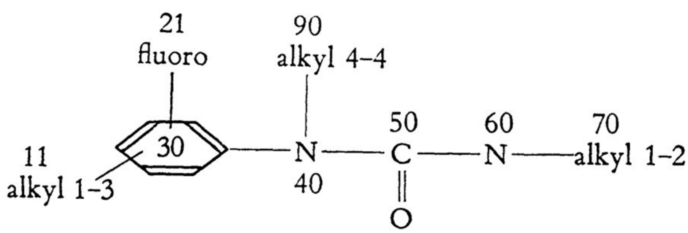
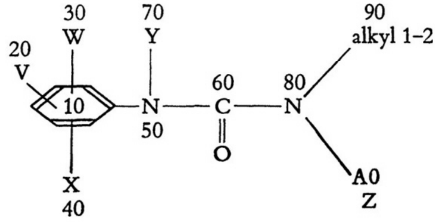


FIGURE 15.

TEST 2

(Question 2 has 1280 unique embodiments, Disclosure 2 has 48 unique embodiments, and this search is the equivalent of 61,440 separate searches.)

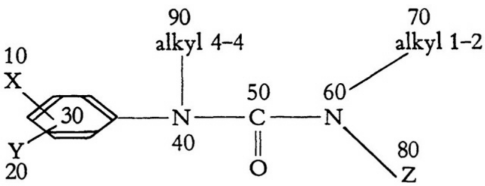
Question 2:



20 23 21 24 22
V = N.S., alkyl 1–3, halogen, or alkyl 1–3 – oxy
30 31
W = N.S. or halogen
40 41
X = N.S. or halogen
70 71
Y = N.S. or alkyl 1–3
A0 A1
Z = N.S. or alkyl 1–4

FIGURE 16.

Disclosure 2:



11 12
X = alkyl 1–3 or NH_2
22 21
Y = OH or fluoro
80 81
Z = N.S. or alkyl 1–2

FIGURE 17.

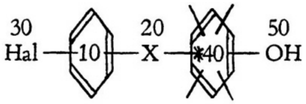
Simulated print-out, test 2: No embodiment found:

Note. This was a "worst case" situation, no answer being possible, and none was found.

TEST 3

(Question 3 has 10 unique embodiments, Disclosure 3 has 8 specific embodiments and this search is the equivalent of 80 separate searches.)

Question 3:

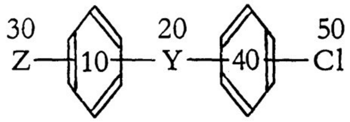


20 21 24 22 25 23
X = N.S., NH, alkyl 1–2 – oxy – alkyl 1–2, or alkyl 1–4

* The unspecified external connections indicate that there may or may not be substituents at these positions.

FIGURE 18.

Disclosure 3:



24 22 21 23
Y = alkyl 2–3 or alkyl 1–1 – oxy – alkyl 1–2
31 32
Z = OH or chloro

FIGURE 19.

Simulated print-out, test 3: Question embodiment:

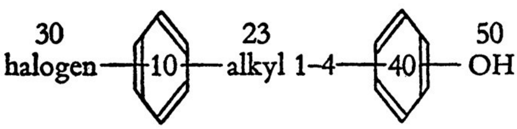


FIGURE 20.

Satisfied by Disclosure embodiment:

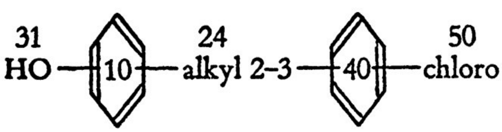


FIGURE 21.

APPENDIX D THE USE OF MACHINES AS AIDS IN THE PREPARATION OF SEARCH FILES

Since the character recognition and mechanical translation arts are not yet sufficiently developed to permit machines to perform the entire job of preparing the search files from printed documents, we must make the best possible use of machines to help human workers perform this exacting and voluminous task. The following is one proposal for a contemplated method of data preparation to be used in the Haystaq system.

A technically trained person will read and analyze each document, note the portions to be encoded, and then prepare a formal abstract or summary of the subject matter. This abstract will indicate the organization of the various levels of disclosure, "index numbers" of the compounds (items) included in the several admixtures (compositions) disclosed, indications of the compositions involved in the various processes, the "accidental" descriptors, and indications of alternativeness, negation, conjunction and other relationships. ("Accidental" descriptors are those characteristics of a material which come to light only because of the environment disclosed for that item in the particular document being treated.) The subject matter of the abstract will then be encoded and transcribed onto punched cards.

Index numbers are unique numbers assigned to each specific compound contained within a collection. Each index number will be represented by an index card (one or more punched cards) which will contain *(a)* synonyms for the compound, *(b)* the index number, *(c)* the structural codes (the types used in 3C and 3D), *(d)* the empirical formula (computed by machine from the structural codes) and *(e)* the fixed descriptors (the types used in 3A and 3B). Such an index card will be prepared the first time that a compound is encountered in preparing the search file.

To compile the complete, encoded disclosure, the machine will record on magnetic tape all codes present on the abstract cards and intersperse in the proper places the codes from the appropriate index cards (except the element-by-element structure code). The tape so prepared will then be processed by the machine according to an "editing" program. This program will generate and insert into the code data for various "mechanical" screens as well as "housekeeping" information for use by the machine in performing the search. Following the editing routine, the encoded data will be checked for mechanical and logical errors by use of another computer program. Since Haystaq employs the element-by-element structure codes on a secondary tape, these codes will be similarly copied from the index cards in the preparation of that tape.

Question data will be similarly prepared.

In order to provide easy access to the desired index cards, an unambiguous filing system must be employed. It is planned to make use of one of the available indexing schemes which generate unique ciphers for each compound (26). Since these ciphers lend themselves to a systematic alphabetization, they can be employed in the manner of tabs on library catalogue card files to enable a machine to locate index cards which contain the detailed codes needed to supplement the abstract in preparing the complete search file.

APPENDIX E SOME COMMENTS ON PARALLEL SEARCHING AND MEMORY DEVICES

The system now being devised for tests on SEAC (and all other systems so far proposed) operates in such a manner that a limited number of questions is held in relatively fixed storage and the entire—very large—collection of permanent data to be searched is serially compared with the internally stored static questions. This is comparable to the FBI's looking for a person whose characteristics they know by making the entire population of the country come to Washington and walk past the FBI building in single file.

A more desirable approach would be to keep the large volume of data to be searched in a static, fixed storage and to use the questions as input. Under such a system, the search would proceed directly from question to location of an answer (if any existed) in the static fixed storage. Questions would be asked serially in rapid succession and the output speed could be made to depend upon the speed with which questions could be propounded as input. This would be of the same order of speed as the input in the system first mentioned above.

By what means could the disclosure file be stored so as to provide truly parallel (i.e., simultaneous) access to all disclosures and permit instantaneous retrieval? Emphasis on memories providing large capacity, rapid access, and rapid retrieval has led to the development of a large volume of factual information concerning the characteristics of a variety of memory systems. More emphasis must be placed on determining the basic characteristics of memory systems in general. Research in this direction could well lead to the birth of systems radically different from any now contemplated. More consideration should be given to deriving systems which can provide unlimited polydimensional access to stored information. Presently available storage devices are relatively bulky in size and most of them are at best two-dimensional in character.

Some work has been done along the lines of polydimensional storage but a more intensive effort in this direction is needed. With the accent on research in solid state physics today, perhaps some investigations into the use of intersecting planes in a crystal, as a memory device, might be in order. Or perhaps some bright young scientist might investigate the use of complex wave forms as storage media and their response to question wave forms by (1) analysis and (2) subset or component wave matching. The area of "self-tracing cross-path systems" should be investigated since such systems offer a possible means for storing and retrieving complex types of internal relationships of disclosures. (See Section III, under "Machines with Learning Ability," part *(c),* for one concept of a self-tracing cross-path system.)

It is desired to de-emphasize neither the importance of volume or cost in storage requirements, nor the speed of access and retrieval, since these are important economic factors. However, the economic aspects of this field should be considered secondary to the principal task of bringing to a higher state of maturity the relatively elementary electronic machines of today as tools for information retrieval and literature searching. The "automatic library" of Bush (22, 23) and other men of vision must be capable of being embodied in some usable form if the results of human investigations in the realm of science are to be made available to further workers in the field, and the concepts mentioned above must be pursued before this end can be attained.

REFERENCES

1. METCALFE, J., *Information Indexing and Subject Cataloguing,* The Scarecrow Press, Inc., New York, N.Y., 1957.
2. NEWMAN, S.M., Problems in Mechanizing the Search in Examining Patent Applications, *Patent Office Research and Development Report No. 3.,* Department of Commerce, Washington, D.C., 1956.
3. LANHAM, B.E., J.LEIBOWITZ, and H.R.KOLLER, Advances in Mechanization of Patent Searching—Chemical Field, *Patent Office Research and Development Report No. 2,* Department of Commerce, Washington, D.C., 1956.
4. LANHAM, B.E., J.LEIBOWITZ, H.R.KOLLER, and H.PFEFFER, Organization of Chemical Disclosures For Mechanized Retrieval, *Patent Office Research and Development Report No. 5,* Department of Commerce, Washington, D.C., 1957.
5. BAR-HILLEL, Y., A Logician's Reaction to Recent Theorizing on Information Search Systems, *American Documentation, 8,* [2], 103–113 (1957).
6. LUHN, H.P., A Statistical Approach to Mechanized Literature Searching, International Business Machines Corporation, Research Center, Poughkeepsie, N.Y., Jan. 20, 1957.
7. PERRY, J.W., and ALLEN KENT, *Documentation and Information Retrieval,* Interscience Publishers, Inc., New York, 1957.
8. PFEFFER, H., H.R.KOLLER, and E.MARDEN, A First Approach to Patent Searching Procedures on Standards Electronic Automatic Computer (SEAC), *Patent Office Research and Development Report No. 10,* Department of Commerce, Washington, D.C., 1958.
9. RAY, L.C., and R.A.KIRSCH, Finding Chemical Records by Digital Computers, *Science, 126,* 814–819 (1957).
10. OPLER, A., and T.R.NORTON, New Speed to Structural Searches, *Chemical and Engineering News, 34,* 2812–2816 (1956).
11. Digital Computer Newsletter, *Journal of the Association for Computing Machinery, 4,* [4], 547–548 (1957).
12. HATTERY, L.H., *Public Administration Review, 17,* [3], 159–163 (1957).
13. LOCKE, W.N., and A.DONALD BOOTH, *Machine Translation of Languages,* The Technology Press of the Massachusetts Institute of Technology, Boston, and John Wiley and Sons, Inc., New York, 1955.
14. WILKES, N.V., Can Machines Think? *Proceedings of the I.R.E., 41,* [10], 1230– 1234 (1953).
15. SHANNON, C.E., Computers and Automata, *Proceedings of the I.R.E., 41,* [10], 1234–1241 (1953).
16. FRIEDBERG, R.M., A Learning Machine: Part I, *IBM Journal of Research and Development, 2,* [1], 2–13 (1958).
17. FAIRTHORNE, R.A., The Patterns of Retrieval, *American Documentation, 7,* [2], 65–70 (1956).
18. NEWMAN, S.M., Linguistics and Information Retrieval: Toward a Solution of the Patent Office Problem, *Monograph Series in Linguistics and Language Studies No. 10.* Georgetown University Press, Washington, D.C., 1957.
19. NEWMAN, S.M., Linguistic Problems in Mechanization of Patent Searching,

Patent Office Research and Development Report No. 9, Department of Commerce, Washington, D.C., 1957.

20. Electronic Computer Study of English Syntax Patterns, *Technical News Bulletin,* National Bureau of Standards, *41,* [6], 84–86 (1957).
21. SHERA, J.H., Research and Developments in Documentation, *Library Trends, 6,* [2], 187–206 (especially p. 202), (1957).
22. BUSH, V., So We May Think, *Atlantic Monthly,* pp. 101–108, July, 1945.
23. BUSH, V., For Man to Know, *Atlantic Monthly, 196,* [2], 29–34 (Aug., 1955).
24. BAILEY, M.F., B.E.LANHAM, and J.LEIBOWITZ, Mechanized Searching in the U.S. Patent Office, *J. of the Patent Office Society, 35,* [7], 566–587 (1953).
25. OPLER, A., Dow Refines Structural Searching, *Chemical and Engineering News, 35,* [33], 92–96, (1957).
26. Which Notation? *Chemical and Engineering News, 33,* [27], 2838–2843 (1955).
27. Seeing—Eye Computer, *Time, 71,* [11], 61, (1958).

A Proposed Information Handling System for a Large Research Organization

W.K.LOWRY and J.C.ALBRECHT

The governing factor in the useful application of information is that determined by need rather than the totality of information available. Information is available in amounts far greater than required by any single research organization and to the degree that it is extraneous to needs it constitutes a barrier to research progress if attempts are made to handle it in an information system. To determine what information is relevant, to locate, acquire, and announce it for ultimate use constitute the preliminaries for effective storage and retrieval. They are no less important than techniques of information control such as cataloging, indexing, and coding, and when properly oriented to needs they assist greatly in the simplification of control requirements by reducing the amount of information to be controlled. The proper control of information is a prerequisite to its effective storage and use and should also be governed by the needs of the user rather than the total possibilities for control offered by a body of information.

There have been innumerable attempts to solve the difficulties of information storage and retrieval with insufficient consideration having been given to the reduction of problems in the preliminary phases of the information handling cycle. The general failure of these attempts demonstrates the necessity for delineating true information requirements, reducing the amount of irrelevant material in an information system and designing systems for efficient operation in the searching procedure. By proper attention to each phase of the information handling cycle it is possible to make good use of presently available electronic equipments to aid in storage and retrieval. Further aid is possible by employing oral techniques to speed up the flow of technical communication within a research organization. This paper will suggest a system of information

W.K.LOWRY and J.C.ALBRECHT Bell Telephone Laboratories, Murray Hill, New Jersey.

handling for a large research group but it is believed that most of the approaches and procedures are equally applicable to small organizations.

Numerous problem areas of information handling contribute to dissatisfaction with most systems presently in use. This is not to say that all current systems are unsatisfactory. There appear to be a number, particularly for scientific disciplines enjoying well developed taxonomies or in limited fields, which work quite well. On the other hand, satisfactory systems for handling large amounts of information to meet the diverse needs of many research organizations still need to be developed. In such situations problems begin when information is first put on paper by a research worker and continue to accumulate through subsequent stages of the communication cycle. The net effect is a serious and complex situation with respect to the latter stages concerned with storage, retrieval and use of information. It is the reduction of this entropy in information systems which must be achieved and the following proposals are governed largely by this objective.

ESTABLISHING THE NEED FOR INFORMATION

In all research organizations information is available from two general sources, that produced by the group itself and that available from outside sources. It is assumed that internally produced information is more reflective of an organization's research activities and interests than that produced externally, and it is recognized that the need for externally produced information is at least as great as that produced internally. From a practical standpoint it would seem desirable as a first step in determining the information requirements of a particular research group to establish these needs in relation to (1) the research projects being pursued, (2) the scientific and technical interests of the group members, and (3) the information produced by the group itself. All these relationships may be established readily since they are subject only to internal considerations and are not affected by outside factors beyond the group's control. Suggested techniques for exploitation of these areas will be described in the following paragraphs. The first objective, it should be noted, is to establish a dictionary or list of information requirements.

In most research organizations it is necessary to prepare a description or justification for each research or development project before funds, personnel, and facilities are authorized. These statements in themselves are frequently sufficiently informative to permit clear definition of areas where information will be needed. They will also usually carry the signatures of those responsible for project work; so questions concerning information requirements can be readily answered. Analysis of work authorizations has the additional advantage of indicating information needs before or at the time project work starts. The search for information, its compilation, and presentation at an early date should reduce both the time and money requirements for research projects.

Successful research administration has recognized the long-range advantages resulting from a policy which permits the individual scientist to explore areas of knowledge which are not task-oriented. The special scientific interests held by members of a research organization sometimes appear to be only remotely related to current research projects but encouragement of such interests has frequently resulted in comprehension of relationships in knowledge which open new frontiers in science having practical applications. The larger the research staff and the more diversified the program, the more likely it is that the aggregate of individual interests will approximate those inherent in the research projects of a laboratory or institute. There are several methods of determining individual interests, one of which is by use of a well-designed questionnaire to be filled in by scientists and engineers. This approach would provide a second possibility for establishing information needs and the preparation of a list of information requirements. In another part of this paper where oral communication techniques will be discussed, reference will be made to the development of a catalog of skills as an aid to internal communication within a group. To determine individual skills, a questionnaire approach is also suggested, so it seems advisable to combine the inquiry on interests and skills in one questionnaire.

A third approach to establishing information requirements is that based on areas of interest as reflected in the publications produced by a research organization. These constitute the written record of scientific activity and represent the distillation of long hours of intellectual effort directed toward research objectives. As such, they provide a firm basis for the determination of information needs pertinent to these objectives. Analysis of internally produced publications for subject content will assist in the preparation of an index of information requirements.

The methods suggested imply participation by the research staff. An accurate estimate of requirements is unlikely if it is made by any other means and the time required of the research staff for this effort is insignificant compared to that required for laboriously searching through irrelevant information unnecessarily controlled by generally unsatisfactory techniques. The principal contribution of the scientist or engineer in this respect is that of indicating areas of pertinent interest in a language generally acceptable to his professional community. The bibliographical aspects necessary to the development of the information system and the general administration of information services should be the responsibility of personnel having experience in these areas.

PREPARATION OF AN INDEX OF INFORMATION REQUIREMENTS

The foregoing indicates three sources for determining the information requirements of a research organization. To specify these requirements in a useful form, it is suggested that they be expressed as a list of scientific and technical terms initiated by members of the research staff to insure a terminology which is familiar to them. This suggests the need to prepare brief instructions for assigning terms to documents which are most likely to meet the direct or related interests of others. The specific rules should be developed to meet the needs of the individual research organization and, when formulated, they should be published, distributed, and considered standard for indexing purposes throughout the organization.

In accordance with the Instructions for Indexing, current work authorizations could then be reviewed by personnel responsible for each research project and a list of terms reflecting the information requirements for a project compiled. The terms assigned by each project group would then be forwarded to an Index Review Panel.

To determine the scientific interests of individual scientists and engineers, a questionnaire should be designed which would explain the objectives of the general indexing program and the method of listing interests in accordance with the indexing instructions. The questionnaire would also provide a section for listing the specific research project interests of each individual if these differ from personal scientific interests. A third section would permit the listing of technical skills which the individual feels competent to discuss with other members of the staff who might be in need of his advice. When completed, these questionnaires would be sent to the Index Review Panel.

To obtain a list of terms pertinent to internal publications, authors would review their papers and submit appropriate terms in accordance with the general Instructions for Indexing. The extent to which previously published reports and papers should be indexed would be determined by the individual research organization. It may be that analysis of publications issued during the previous year would be sufficient to establish the initial list of terms or it could extend backwards for five to ten years. Terms assigned to internal publications would be forwarded to the Index Review Panel.

After receiving suggested terms from the research staff, the Index Review Panel would review them for form and to eliminate inconsistency, synonymous terms, etc., and would prepare a basic list and a list of cross references. The cross-reference list would include all terms submitted but not included in the basic list. The Index Review Panel should be comprised of scientific and technical personnel having backgrounds representative of the major areas of research interest and should be headed by a person having the necessary training and experience in the preparation of indexes. It is assumed that the special expertise required for the development of an effective index would not be available from research personnel and that their technical contribution would have to be strongly supported by a competent indexing specialist.

After compilation of the master List of Index Terms, it would be published and distributed to all scientific and technical personnel for use both as an aid in subsequent indexing and as a guide for terms to be used in requesting information at a later date. New terms would be added to the list only after review by the Index Review Panel and revisions of the published list would be issued as determined by the changing interests in the research program.

Since the List of Index Terms would reflect the scientific interests of the research staff, it would also serve as an aid in selecting publications prepared externally which would be of value to internal information needs. As such, it becomes a useful guide for the acquisition of information pertinent but not extraneous to those needs.

ESTABLISHING INFORMATION CONTROLS AT THE TIME OF PUBLICATION

As was noted earlier, information handling problems begin with the preparation of new information by a research worker. Certain controls can be exercised at the time of publication which will facilitate subsequent indexing, abstracting and announcement, and for this purpose Instructions for Abstracting and an Information Control Sheet are useful. Rules for preparing author abstracts should be developed and distributed to all technical personnel in the research organization. Advice concerning the need to be informative and the necessity to avoid the use of terminology, code words and abbreviations that are not well known outside a particular area will do much to aid in communicating information within an organization. A control sheet utilizes the advantages of standardized presentation of information for indexing and abstracting and can serve as final copy for the preparation of abstract bulletins. The space provided on the control sheet for the assignment of index terms will facilitate the prompt preparation of subject indexes for abstract bulletins.

The information control sheets would be forwarded to the Index Review Panel for an additional purpose. Usually an author is well qualified to suggest terms reflecting the technical content of his paper, but he may not recognize the usefulness it holds for other scientific or technical fields. These relationships are important in view of the interplay of interests between various disciplines. The review panel, representing a cross section of total interests, can assist materially in recognizing them and assigning additional useful terms.

CORRECTING SUPERFLUITY AND PAUCITY IN INFORMATION ANNOUNCEMENT

The large amount of information published has resulted in a common complaint that scientists are unable to determine what is pertinent to their interests. This is supported by the constantly increasing number of scientific and technical journals, the delays and duplication associated with abstracting services, and the time required to sift through the chaff of newly published materials to get a few useful kernels of information. This situation results in a rate of information loss which probably approaches that of information use and points up the need for improved techniques of announcement to individual scientists. This is not only true for scientific publication generally, but frequently applies to publications prepared within large research organizations. Previous suggestions on relating information needs to research objectives also apply in respect to announcement of information although the direction of information flow is reversed. In this phase of information handling, the needs of the research group and its individual members have been established as noted earlier, and the present concern is that of further reducing the amount of technical man-hours required for literature review by developing a system of announcement specifically tailored to individual requirements in addition to the announcement system used to cover the needs of the research group as a whole.

To meet overall needs, an announcement bulletin issued regularly and containing abstracts of all internally produced publications can be prepared readily from copy supplied by authors on the Information Control Sheet previously mentioned. The standard format of the control sheet permits ready preparation of the bulletin by using xerography and offset processes. Abstracts or titles of externally produced journal articles and reports selected and indexed by the Index Review Panel could be included in the same bulletin after copy has been prepared for these items. By assigning serial numbers in sequence to each item, electronic searching and location of desired items is possible. Serial numbering does not preclude the possibility of a classified arrangement within the bulletin, if this is desired, since numbers may be assigned after page format is completed. An index may be prepared for each issue and cumulatively if desired by use of the Flexowriter programatic automatic writing machine with its tape-to-card punch feature or by any of several other techniques.

In addition to the general announcement bulletin, it is also possible to prepare announcement lists pertaining to the specific interests of individual scientists and engineers as indicated by their earlier statements of information requirements. By utilizing the electronic techniques to be described later, the code numbers representative of fields of interest for each member of the staff can be matched in search with those relating to the indexed documents and pertinent document numbers obtained. These special interest reference lists could be attached to the abstract bulletin before mailing and would serve to pin-point items of particular significance to the recipient. The time required to print out such lists will depend upon the number of matches obtained and the output device used but the time for search can be estimated fairly accurately. Using punched cards which specify terms of interest for each of 2000 individuals and assuming that each month 2000 documents had 30 index terms assigned as an average, there would be on the order of 30 feet of magnetic tape required to store the term and administrative codes for the documents. With an automatic card feed and a continuous tape, the time required for search and solution of the problems of 2000 individuals would be of the order of one hour. This assumes the use of buffer output tapes and off-the-line print out of the announcement lists which would take from two to seven hours, assuming a relatively inexpensive line printer. The automatic preparation of 2000 monthly lists by this technique would aid in meeting the requirement for prompt announcement of information of special interest to individual scientists. The details of this procedure will be explained later.

ORAL COMMUNICATION IN THE INFORMATION PROCESS

One of the consequences resulting from the large amount of information which is inaccessible is the increasing reliance which scientists place upon other scientists for their information needs. The increase in oral communication to meet information requirements is undoubtedly influenced by the failure of conventional library techniques to be of sufficient assistance to the scientific community in finding and making available information when it is needed. The procedure of obtaining knowledge from experts in various fields is in itself becoming more difficult and in a large organization presents a number of problems similar to those encountered in exploiting written information. Personal contact becomes difficult when a research organization has its facilities dispersed, its program diversified, and its staff enlarged to the point where it is difficult to determine just who is the expert to consult. Another factor contributing to the increasing failure of the personal contact technique is the relatively high turnover of personnel which most large organizations experience. The ability to communicate personally is particularly difficult for newly employed scientists and engineers who are unfamiliar with the talents which their new organization may possess. Even with these barriers to personal communication, it remains one of the most used methods to obtain information and in view of this efforts should be made to improve the possibility of personal contact in large groups.

As noted earlier, the questionnaire to be filled in by all members of the research staff provided a section where each person would indicate the special areas in which he felt competent to give advice. The preparation of a Register of Skills, based on information from the questionnaire, would be a relatively simple task and would do much to facilitate personal communication. This register could be maintained in a central place, its availability made known generally, and when an individual wished to discuss a particular area of interest, the names and telephone numbers of qualified co-workers could be provided. The method of establishing the register, maintaining it, and providing service on it could be completely manual or it could be semi-automatic by using edge-punched cards or any of several other suitable approaches. It is entirely possible that certain staff members may indicate proficiency in a particular area and that their estimate of personal ability could be overemphasized; therefore, it would perhaps be useful to have their immediate superiors review these estimates. There are, of course, other approaches to establishing the Register of Skills, such as review of statements of experience and training which are usually available from personnel offices or an approach could be made directly to the supervisory level for an opinion of the skills available from persons within their department. It is believed, however, that the suggested method has advantages not held by other methods.

By way of supporting the foregoing suggestion, a survey was made in at least one large research organization to determine what sources of information were most used. The results of this survey indicated that in the case of about 80% of the individuals surveyed, their answer was that "they asked somebody." In about 50% of the cases (the second most used source), internal publications produced by members of the organization were found most useful. Further support for improving oral communication is evident in a recent editorial by Edmund C.Berkeley appearing in *Computers and Automation,* in which he states

…there ought to be much better techniques of communication at computer meetings: (1) The tag which you wear at a meeting should show: your name, your organization, *your main interests*. This would aid communication by telling whether or not this man is interesting to me; we should talk together.

Electronic storage, search, and retrieval of information

Intelligent selection of material to meet user requirements as outlined in the preceding suggestions will substantially reduce the amount of information which must be stored for future reference. However, in a large organization the amount of information stored will soon exceed the capacity of manual or commonly accepted retrieval methods. A device having the intelligence of the human mind and the speed of a modern computer would be an ideal solution to the problem of data retrieval since the mind is extremely flexible and is capable of subtle associations and correlations necessary to efficient retrieval. Unfortunately, however, such intelligence is difficult, if not impossible, to achieve fully in a machine.

Where material to be retrieved can be clearly specified as in the case of a parts man searching for a particular component, simple machine methods are acceptable since the problem can be clearly outlined in a search request. A more difficult problem, however, lies in the retrieval of information relating to desired subject matter, such as is the case in the compilation of a bibliography or in the situation where a scientist is interested in information pertinent to specific subject matter. In these situations, the problem is extremely complex since the subject matter of interest may be discussed in detail in certain of the information indexed or merely just in passing in other material, depending upon the scope and setting of the information source. For example, a development engineer interested in a specific electric circuit or mechanical movement faces the problem of finding not only all information directed specifically to the area of his interest, but also all information wherein his particular area of interest is treated in context. In most instances, the area of interest of a person seeking retrieval of information is not fully matched by the scope and setting of information stored and indexed for future reference. It is in this situation that the culling process achieved through the associations and correlations of the mind may be employed to advantage to determine what stored information may be of interest, even though it does not fully define the user's area of interest.

A simple technique which approaches a function of the mind and which may be readily embodied in machine searching of information files comprises the establishment of two classes of indexing terms, namely general and essential. From this classification, it is not to be inferred that an indexing term is defined in the standard list of terms as being general or essential, but rather is so defined at the time a particular search is undertaken. Indexing term, as employed herein, relates not only to single keywords, but also to indexing phrases, names, etc.

This classification of terms in data retrieval is not limited to a specific vocabulary or hierarchy of definitions, but can be employed to advantage regardless of the indexing approach. As previously discussed, all material stored for future reference is indexed in accordance with a standard list of terms. In requesting a search, the subject matter about which information is desired is defined in accordance with terms drawn from the standard indexing list. The person requesting the search or possibly a skilled operator determines which of these terms must apply to stored information for it to be of interest, and therefore are essential to the search, and which other terms should, but need not apply, for the document to be of interest, and are therefore termed general terms. A direct approach to the problem of retrieving information in accordance with a request indicating essential and general terms is to record the identity of all documents to which the essential terms and, in addition, a threshold number of general terms apply. In accordance with such a system, any or none of the search terms may be designated as essential, and the number and order of search terms is not limited by the action of the machine.

This philosophy of data retrieval can probably best be understood by reference to Figs. 1–5.

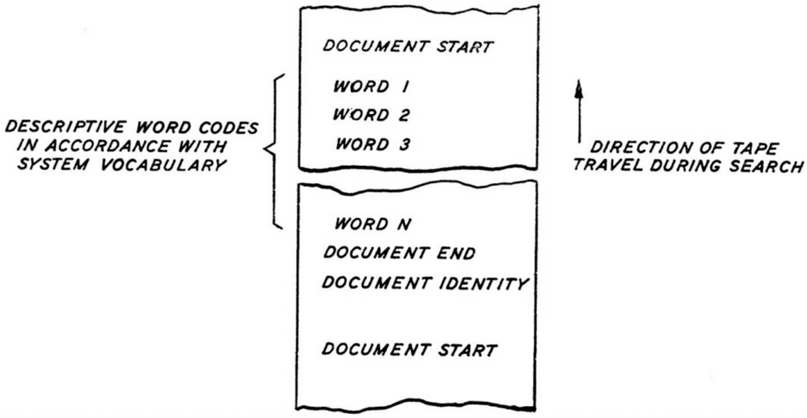


FIG. 1. The arrangement of administrative and indexing terms in a magnetic tape employed in the proposed system.

In Fig. 1, the direction of tape travel is as indicated. The document "Start" and document "End" codes are special administrative codes which will be discussed later in detail, the words 1 through N are codes representative of indexing terms taken from a standard list or vocabulary previously described, and the document identity is, as might be expected, a code representative of the document to which the preceding indexing codes apply.

The operation of the proposed search machine can be understood by a description of Fig. 3 in terms of a particular search problem.

A development engineer faced with the problem of designing a new high-frequency and extremely stable oscillator might request a machine search to determine the identity of available written material relating to his particular design problem. Figure 2 shows a suggested form whereby the engineer may request a machine search to provide him with the desired information. In Fig. 2, only five identifier words are shown associated with the search. However, this is not a limiting number, and any reasonable number of terms within the capacity of the machine could be specified. The identifier terms are taken from the previously described standard vocabulary or list of indexing terms, and the identifier code representative of the identifier words are similarly obtained from the standard vocabulary. The translation from identifier word to identifier code has been shown as a manual translation; however, at the expense of added cost and machine complexity, a machine translation could be performed if the human translation were too burdensome. In the right-hand column of the request form of Fig. 2, a note is made as to whether each of the identifier words is of general or essential interest to the particular search.

REQUEST FOR MACHINE SEARCH

WORD	IDENTIFIER WORDS	IDENTIFIER CODE	ESSEN-TIAL
1	TRANSISTOR	1011	NO
2	OSCILLATOR	2369	YES
3	MODULATED	4047	NO
4	RADIO FREQUENCY	6897	NO
5	CRYSTAL CONTROLLED	1645	NO

REQUESTED BY	WHEN REQUIRED	CHARGE TO	NUMBER OF GENERAL IDENTIFIERS REQ. TO MATCH
JCA	5-15-57	18373	3

FIG. 2. A simple example of an office form whereby a request for machine search may be made.

In the example, the word oscillator has been marked as essential, while the words transistor, modulated, radio frequency, and crystal controlled are marked as being nonessential or general terms.

With the exception of the box labeled "Number of General Identifiers Required to Match," the remainder of the information at the bottom of the request form is for administrative purposes. The notation "3" under the lower right-hand box of Fig. 2 indicates that the person requesting the search is interested in all documents relating to oscillators and to which three or more of the general terms apply. For example, documents relating to radio-frequency, crystal-controlled, modulated oscillators, as well as documents relating to transistor, radio-frequency, crystal-controlled oscillators, are of possible interest. Further, all documents to which the word oscillator and to which, in addition, any combination of three or more of the general terms applies are of possible interest.

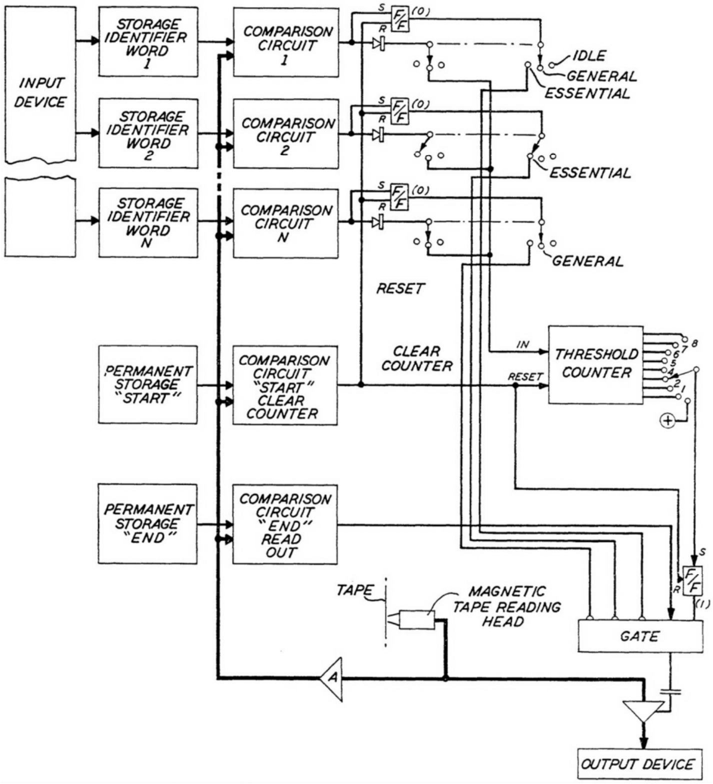


FIG. 3. A basic block diagram of a special purpose computer arranged to retrieve data in accordance with the previously described search approach.

Figure 3 shows in block diagram form the essential elements of a machine capable of solving the problem, established in the previously described request form. The input devices are shown in general form as these may range from a simple arrangement of a plurality of keys or switches manually settable to establish the codes representative of the indexing terms or to more complex arrangements, such as punched cards, magnetic tape, or operator key sets which when manipulated are effective to generate codes representative of the indexing terms.

In this discussion a specific code for storing information in the search files has not been designated, as this decision is relatively unimportant to the operation of the machine so long as an efficient unambiguous coding system is employed. A decision as to what code would be most appropriate depends in part on the codes used by other machines employed in conjunction with the retrieval machine, and such factors as the addition of error detection or error correction would in large depend upon the accuracy demanded of the system.

The storage blocks for identifier words 1 through N are shown as separate from the input devices. However, in the case of simple input devices comprising keys or switches or even punched cards, the storage would be included in the input device rather than as a separate box as shown. The two permanent storage elements marked “Start” and “End” are merely wired arrangements in which codes representative of the previously mentioned documents “Start” and documents “End” codes are stored. The comparison circuits 1 through N and the comparison circuits “Start” and “End” are arranged to compare information stored in an associated word storage and parallel information read from the magnetic tape, and to provide an output pulse whenever coincidence occurs.

In our example, the identifier words of the request, by means of the input devices, are entered in any of the descriptive word storage blocks 1 through N without regard for order. A double deck switch such as is shown schematically at the right side of Fig. 3 is associated with each of the comparison circuits 1 through N. The purpose of these switches is to designate whether the identifier stored in the associated storage block is of general or essential interest to the search, or if the particular associated storage block is not in use, to establish the idle condition with regard to that particular comparison and storage circuit.

The magnetic tape has a plurality of tracks; the exact number of tracks is not at this point essential to an understanding of the proposed system. This number would be determined by the number of identifier words employed in the system and by the coding scheme employed to identify the documents. For example, if a continuous file covering a number of years were employed, the number of tracks would have to be sufficient to record a code descriptive of the largest anticipated number of documents in the system. However, if informa tion is stored on annual tapes, considerably fewer tracks would suffice, since redundant information defining the year can be omitted.

The magnetic tape, arranged as shown in Fig. 1, travels past the parallel magnetic tape reading heads, as indicated in Fig. 3, to provide signals representative of the information stored in the tape, to the line amplifier and to the gated output amplifier. If it is assumed that the magnetic reading heads first encounter a document "Start" code, the information from the line amplifier and that in the "Start" permanent storage will match, and a "Clear Counter" signal will be generated in the "Start" comparison circuit. The clear counter signal resets the threshold counter to "O" count and resets the flip-flop inputs to the output gate. The machine is now prepared to examine the document identifiers contained in the tape between the just encountered document "Start" code and the document "End" code succeeding word N. As the document identifier words are read serially from the storage tape, the code representative of each word is compared in parallel with the code entered in the storage blocks at the time the search is initiated. If a match is found, an output is provided by the particular comparison circuit. For example, if the code representative of the word "transistor" is stored in the first storage block, the first comparison circuit output conductor would be energized each time the code representative of "transistor" is encountered in the search file. The setting of the rotary switch associated with a particular comparison circuit determines whether the output signal from that comparison circuit will be directed to the input of the threshold counter which is associated with the general indentifier terms of the search or to the output *and* gate through a flip-flop circuit as is the case when the term is stated to be of essential interest to the search. In Fig. 3, the rotary switch associated with the first comparison circuit is shown in the general position, and thereby the output of the first comparison circuit is connected to the input of the threshold counter. Accordingly, each time the word stored in the first word storage block matches the code read from the magnetic tape, the threshold counter will be advanced one count. The flip-flop associated with the first comparison circuit is set whenever a comparison is found. However, with the rotary switch in either the idle or general position, the output of the flip-flop will be open-circuited. When the rotary switch is in the essential position as is the case of the switch associated with the second comparison circuit, the output of the associated flip-flop will be connected to an inhibit lead of the output gate. Resetting of the flip-flop energizes its output conductor, thereby inhibiting the output gate until such time as the comparison is found for the essential identifier which effects setting of the flip-flop.

If the information encountered between a document "Start" and a document "End" code matches the conditions imposed by the search problem, the identity of the pertinent document should be recorded in the output device. The flip-flop shown adjacent to the output gate is controlled by the threshold counter output which, in accordance with the search request, is set in position 3. Accordingly, if the threshold counter reaches a count of 3 between a document "Start" code and a document "End" code, an output signal is provided from the threshold counter to set the associated flip-flop to its "1" state. If the flip-flops associated with essential words and the flip-flop associated with the threshold counter are all set at the time the document "End" code is read from the magnetic tape files, the output gate will be enabled to provide an output pulse to in turn enable the output amplifier. Accordingly, the information immediately succeeding the document "End" code, namely the document "Identity" code, is gated to the output device. The transient signal from the output gate is employed, therefore only the document "Identity" is read, and additional extraneous following information such as the document "Start" code is not read out.

As in the case of the input devices, the output devices are shown only in general form. Obviously a very high-speed direct printing device would suffice for this application. However, on-line use of such a device probably would not be economically sound. Accordingly, in most instances, buffer storage of one type or another would appear to be advantageous. If a rule of operation is established such that a minimum number of lines on the search tape will be left between identity codes, a simple constant-speed magnetic tape could be employed to advantage as an output buffer. For example, if a minimum of thirty lines were reserved for each document indexed, the output tape could be run at a constant speed of 1/30th of the speed of the search tape. Accordingly, identities read from adjacent entries on the input tape will appear as adjacent entries on the output tape. Where a large number of documents exist in the file between the identities of pertinent references, spaces will occur between identity numbers in the output tape. This does not appear to be a serious problem since the output tape may be removed and employed to control printing devices away from the operation of the search machine in a typical off-the-line mode of operation.

The subject system readily lends itself to an orderly increase in search capacity. An extension of the system of Fig. 3 to permit three simultaneous searches is shown in Fig. 4. An orderly expansion is possible since the identifier word storage and comparison circuits are not permanently associated with a particular search problem, but rather are associated with a particular problem in accordance with search requirements. For example, if thirty-word storage and comparison circuits with associated flip-flops were provided to operate with three search circuits each comprising a threshold counter, a flip-flop, an output

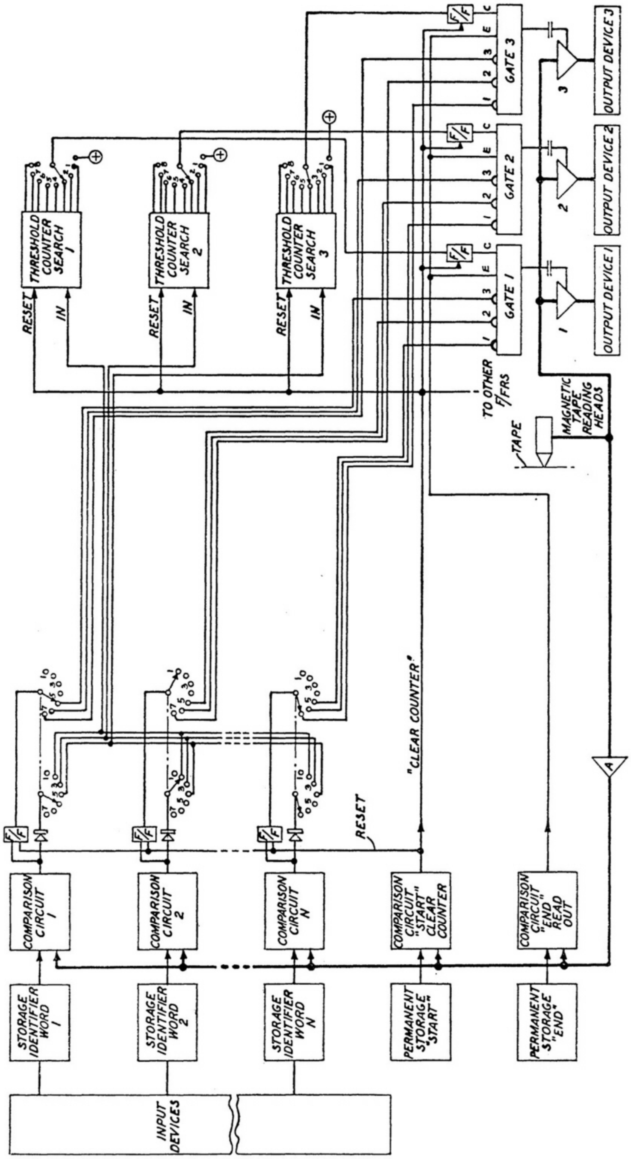


FIG. 4. A block diagram showing an expansion of the arrangements of Fig. 3 to permit simultaneous solution of more than one search problem.

gate, a gated output amplifier, and an output device, it is doubtful that each problem would always require one-third of the comparison and storage circuits. It appears more likely that certain search problems would require more than ten comparison circuits while others would require less; therefore, the flexibility available through the use of the rotary switches which direct the output of the comparison circuits to their proper search tasks is of considerable worth. In Fig. 4, identifier storage and comparison circuits are assigned to a particular search at the time a search is initiated. It is possible that identifier codes representative of the same indexing term may be entered in separate storage circuits which are associated with separate searches. Accordingly, it is a distinct possibility that comparison circuits not associated with the same search will be simultaneously energized to advance the respective threshold counters or to set associated flip-flops.

The operation of the circuit of Fig. 4 is identical to that of Fig. 3 except that now a plurality of searches may be simultaneously undertaken. The separate searches may refer to separate search problems or they may specify different ways of posing a search problem relating to a particular piece of information. Accordingly, the second and third searches may specify only different degrees of search requirements with regard to a single search problem or they may advantageously relate to separate and distinct search problems.

A MECHANIZED SYSTEM OF INFORMATION ANNOUNCEMENT

A monthly notice to pinpoint abstracts of particular interest to personnel of an organization has been mentioned herein without specific reference to machine arrangements. A modification of the arrangements of Fig. 4 to accomplish this task is shown in Fig. 5. It has been estimated that in a large organization this service might be furnished to several thousand individuals each having as many as nine or more areas of interest either relating directly to their present work or to special fields of interest not necessarily related to their present jobs. It has been further estimated that the number of documents indexed each month might run in the order to 2000. A large organization having 2000 scientists each having nine areas of interest would require 18,000 search problems per month merely to prepare the notices to accompany the abstract books. Although 18,000 serial searches of a file having 2000 document entries might appear to be a tremendously lengthy process, the arrangements of Fig. 5 appear to be capable of performing this task in an extremely reasonable period of time.

Figure 5 is an extension of Fig. 4 with specific suggestions in the area of input and output devices; the use of electronic switching means to gate the information signals throughout the system rather than mechanical switches; and the

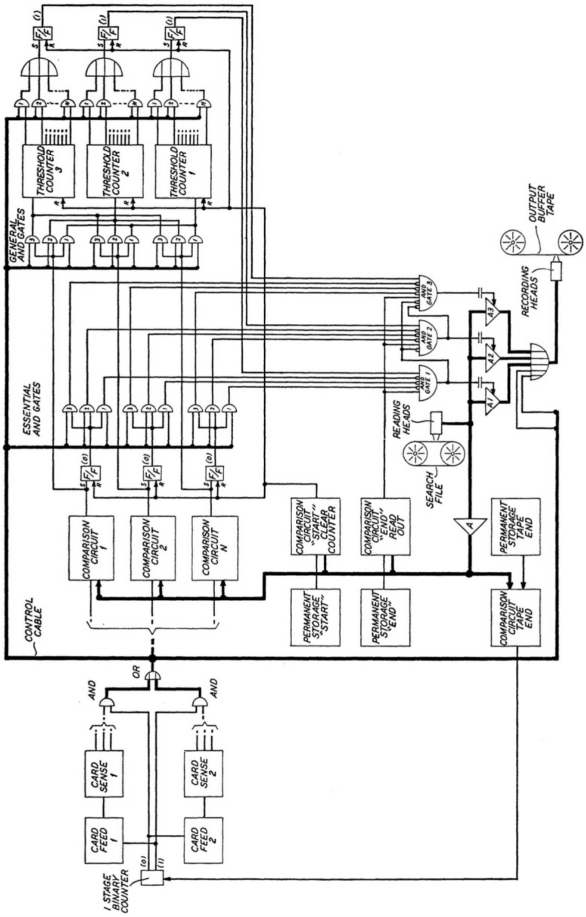


FIG. 5. A variation of the arrangements of Fig. 4 to provide rapid preparation of announcement lists.

use of a continuous tape search file as opposed to the arrangements of Fig. 4 wherein tape is shuttled from reel to reel as the search progresses.

The format of the search file as shown in Fig. 6 is identical to the search file shown in Fig. 1 with the exception of the addition of an administrative code to indicate the end of the indexing information. In Fig. 6, this additional administrative code has been labeled "Tape End."

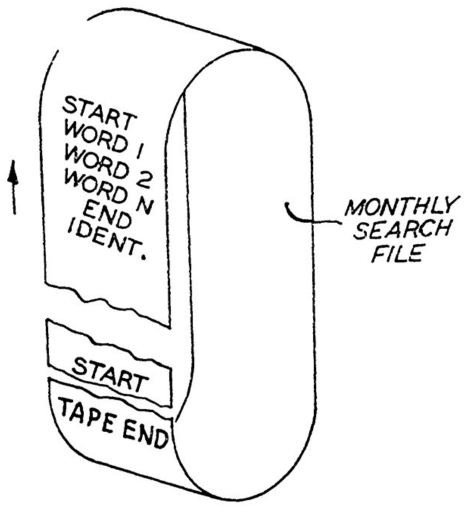


FIG. 6. The search file.

A minimum number of lines are left between identity codes in the monthly search file so that the output buffer tape may be run at a uniform relatively slow speed. As previously discussed, the speed of the output buffer tape would be equal to the speed of the search tape divided by the number of lines between document "Identity" codes.

Information is set into the arrangements of Fig. 6 by means of punched cards which are fed alternately from two piles of cards. The use of two card feed mechanisms considerably relaxes the requirements for such devices with regard to speed of operation. The philosophy of operation of Fig. 5 is identical to that of Fig. 4, but a short description of the overall operation will readily clarify the areas wherein equipment changes have been made.

Assuming as a starting point that there are no cards in the card sensing mechanisms shown at the left side of Fig. 5 and that the search file tape is being run past the reading heads in the direction shown in Fig. 6, the "Tape End" comparison circuit will be energized when the "Tape End" code is encountered in the search file. The comparison circuit output signal is connected to the single stage binary counter and, when energized, will drive the counter either from "0" to "1" or from "1" to "0," depending upon the prior state of the counter. If the counter is driven from "0" to "1," an enabling pulse will be transmitted to card feed 1 to shift a card into the first card sensing mechanism and to energize the *and* output gate associated with the second card sensing mechanism. Accordingly, since there is no card in the second sensing mechanism, there is no possibility of conducting a search during the first pass of the tape file. However, the next time the "Tape End" signal is encountered, the binary counter will be shifted from the "1" to "0" state to enable the second card feed to place a card in the second card sensing mechanism, and to enable the *and* output gate associated with the first card sensing mechanism. Enablement of the *and* gate establishes the desired search conditions within the machine and provides a momentary path from the card sensing mechanism to the recording heads. Accordingly, as soon as a card is sensed, the identity code of the person to whom the information should be directed is entered in the buffer output tape. In this way, document identities read to the buffer tape from the search files in accordance with the search problems defined by the input cards will follow the identity of the person requesting the search.

The input card contains not only the identity of the person requesting the search, but also all information relating to the search such as the codes of the words defining the search problem; the determination of whether these words are general or essential identifiers; the number of the search problem to which these search terms are to apply; and the threshold number of general identifiers which must apply for a document to be of interest.

Figure 5 has been simplified considerably through the use of heavy lines indicating a control cable having a large number of control conductors and a second heavy line indicating a multiconductor bus bar type of interconnection conveying the parallel bits representative of the codes read from the search file. Further, where pluralities of paths exist, amplifiers, *and* gates, and *or* gates have all been shown as represented by a single gate. However, it must be fully understood that in such instances a plurality of gates or amplifiers are necessary.

The search identifying terms sensed from the input card provides parallel inputs to the comparison circuits 1 through N and control information read from the input cards, enables specific *and* gates to establish paths from the comparison circuit output conductors, and their associated flip-flops to either the threshold counters associated with the desired search, or to the output *and* gate associated with the search. Energizing of a comparison circuit output conductor will effect setting of its associated flip-flop, and if a term is stated in the problem to be an essential identifier, one of the gates connected to the "0" output conductor of the flip-flop will be enabled to complete a path from the flip-flop "0" terminal to the inhibit input terminal of one of the three output *and* gates, depending upon the problem to which the comparison circuit is assigned. If the identifier assigned to the comparison circuit is a general term, the flip-flop "0" conductor output will be open-circuited since none of the essential *and* gates associated with that circuit will be energized, and the comparison circuit output conductor will be connected through one of the general *and* gates to the threshold counter of the proper search problem. Accordingly, the selective enabling of the essential and general *and* gates accomplishes the functions previously delegated to the rotary switches of Fig. 4.

The output *and* gates are arranged so that if the "0" output conductor of a flip-flop associated with the comparison circuit is energized and this state is transmitted to one of the three output *and* gates, the output gate will be inhibited. When a comparison is found for an essential identifier, the flip-flop will be set to its "1" state to deenergize the "0" conductor and thereby remove the inhibiting signal from the associated output *and* gates. As in Fig. 4, each time a comparison circuit associated with a general identifier is energized, the threshold counter associated with that search will be advanced one count. The threshold count assigned to a particular search problem is established by enabling one of the threshold counter output *and* gates so that if a counter reaches the preassigned threshold count, the counter output flip-flop will be energized.

Whenever the conditions of a search problem are met between a document "Start" and a document "End" code, the output *and* gate associated with that particular problem will be enabled to in turn enable the associated gated output amplifier. Thereby, the identity of the document meeting the search requirements will be read to the output buffer tape through the recording heads. There is the possibility that a document will be of interest to more than one of the three simultaneous search problems. Since all three searches relate to one man's interest, the output *and* gate may be arranged in a simple preference chain to prevent mutilation of output signals.

After the tape file has been completely searched, the "Tape End" signal will again be encountered to advance a new card into the just used card sensing mechanism, and to enable the output *and* gate associated with the other card sensing device.

The searching process will continue until such time as all cards have been searched. It is contemplated that the output tape would be taken away from the searching machine, and the information thereon printed out in standard off-the-line printing procedures.

SUMMARY

The suggestions made in this paper are directed to several major problems in information handling. They are based on the belief that in most large research organizations these problems exist because sufficient attention has not been

given to systematic study of ways and means to correct them. It is believed that research organizations frequently acquire more information than they need because of failure to identify information needs and that the general philosophy that information needs will be met simply by increased acquisition cannot be supported from the standpoint of economy or practical use of much of the information which is acquired. In this respect, the number of articles selected by the Index Review Panel from journals received could well serve to indicate the value of a particular journal to the organization and should influence decisions on further subscription to little used journals. In broader scope, the determination of information requirements should provide more accurate selection of all forms of information according to stated needs. Three approaches to establishing information requirements have been suggested, the first being concerned with needs in relation to immediate tasks, the second in relation to developing the scientific potential of the staff, and the third in relation to organization interests as reflected in the publication of information resulting from research activities. There are, of course, many other sources which could be utilized to relate information needs to objectives including patent activities, administrative correspondence, laboratory notebooks, and conferences, all of which relate to the three approaches discussed.

In the review and control of information, a primary consideration is the need for participation by the research staff and by scientifically or technically trained personnel in the information group. To the degree that a research organization wants good information control, it will support staff participation in gaining that control. The use of standardized techniques in the information control process will facilitate the orderly and consistent development of a system for effective control. The techniques suggested for announcing new information to members of the staff provide for announcement to the staff generally and to individuals specifically. This approach does not deprive information from the man who enjoys going through the published literature assiduously, but at the same time it does pinpoint items of particular significance to those who are not inclined to do this. In several respects, oral and machine techniques can serve a useful purpose in improving an information handling system, and suggestions for their use have been made. It is not suggested that mechanization is desirable if it is more economical to use manual techniques or when machines are incapable of doing work where human effort is required. An attempt has been made to bring into combination manual, intellectual and machine operations where it seemed advantageous to do so.

Information Handling in a Large Information System

P.R.P.CLARIDGE

The work described[1] in this paper was undertaken in response to a need felt at the Low Temperature Research Station for a collection of information on chemical compounds found in edible plants. Enquiry showed that there would be widespread interest in a collection of this type. That the need was generally felt was confirmed by a reference made to the lack of collected information by the Development Commission (a).[2] The last comprehensive publications of this type (b), (c) were published about 25 years ago, and many of the later publications contain little information of this kind.

Within the last few years, newer methods of analysis, such as chromatography, have facilitated the separation and identification of individual chemical compounds. In consequence, the volume of data being published is showing a rapid increase. In 1955 for example, about 2000 papers reporting definite chemical compounds (as distinct from more indefinite substances such as starch) in flowering plants were abstracted in *Chemical Abstracts* and *Biological Abstracts.*

Some of the papers [e.g., (d)] contain data for a number of plants. On the basis of this sample it was estimated that there would be of the order of 7 million entries (one chemical compound in one plant) in a collection of all the published data on the subject and that the annual rate of addition would exceed 100,000. If the definition of "plant" was to be widened to cover the whole vegetable kingdom and to include the lower plants (e.g., bacteria, yeasts, fungi), the collection would be even much larger than this. It was considered desirable to organize the collection so that this extension would be possible.

The collection was required not only to provide information on named

P.R.P.CLARIDGE Low Temperature Station for Research in Biochemistry and Biophysics, Department of Scientific and Industrial Research, University of Cambridge, Cambridge, England.

[1] The work described in this paper was carried out as part of the programme of the Low Temperature Research Station of the Department of Scientific and Industrial Research.

[2] Lower-case letters in parentheses are references.

compounds in named plants, but also for various generic searches, such as for all compounds having specified groupings in common, and for partial specifications. Correlations between factors such as that between chemical composition and palatability were important. For this purpose the type of indexing to be used would have to be very detailed and constructed in such a way that the maximum amount of information could be displayed on the relations between the entries. A list was made of the various headings under which it was thought information should be indexed, and this was tested on a sample of approximately 2000 entries. The headings were revised (Fig. 1) on the basis of this trial. Subsequent smaller trials have suggested further improvements. There are other items which could advantageously be recorded even though they are not used for indexing the entry, such as language of original paper and type of study (e.g., experimental, comparative). A list of headings including such extra entries and a more adequate selection of types of source has been prepared for use in a larger scale trial of the whole system.

Economic retrieval of all the relevant information in the collection, in answer to an enquiry, was an essential requirement. The ability to select entries showing relationships not suspected at the time of entry of the data was desirable. No system can give out more information than has been put into it, so that the indexing system had to be such that as many implicit relationships as possible would be included when an entry was made. With these considerations in mind, a survey was made of possible indexing systems.

An alphabetical arrangement using plain language entries obviates the need for code books and provides implicit relationships through related meanings of the words used. Urquhart has shown (e) that information in alphabetical subject indexes can become “lost” in the sense that it is not retrieved by a searcher using the subject headings and guides. In his study more of the references looked for could be found by means of the author indexes than were retrieved by a subject approach. In this collection, not only would there be a large number of items to be indexed, but each of these items would have to be indexed from many aspects and at a number of levels of generality. Even the Index to *Chemical Abstracts* does not attempt to provide this last facility to any great extent; it is often necessary particularly for non-chemical entries to look up each member of the class when making a generic search. In an alphabetical subject index also it is difficult to search for information demanded under a partial specification and it was concluded that an alphabetical subject index to a collection of this nature and size was impracticable.

The need for generic searches suggested a classified system. Any form of generic relationship can be taken as the basis of the classification, but once this is chosen, generic searches on other bases have become impossible.

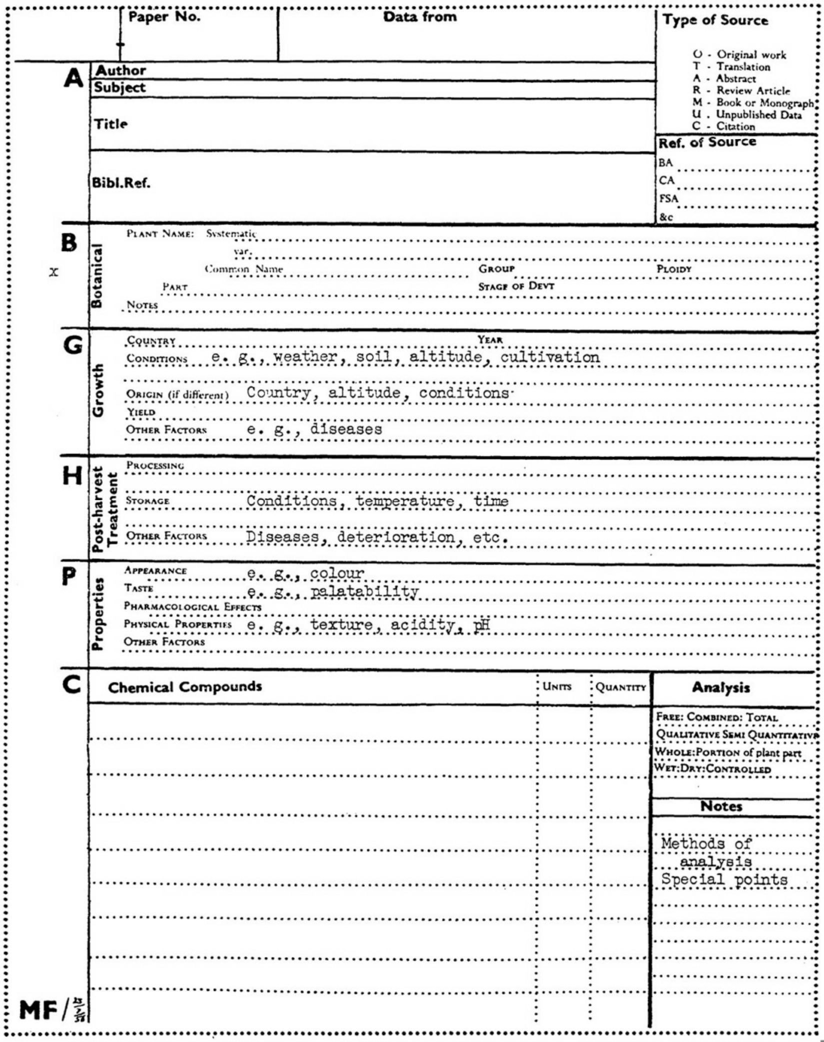

Paper No.
Data from
Type of Source
O - Original work
T - Translation
A - Abstract
R - Review Article
M - Book or Monograph
U - Unpublished Data
C - Citation
A
Author
Subject
Title
Bibl.Ref.
Ref. of Source
BA
CA
FSA
&c
B Botanical
x
PLANT NAME: Systematic
var.
Common Name
GROUP
PLOIDY
PART
STAGE OF DEVT
NOTES
G Growth
COUNTRY
YEAR
CONDITIONS e. g., weather, soil, altitude, cultivation
ORIGIN (if different) Country, altitude, conditions
YIELD
OTHER FACTORS e. g., diseases
H Post-harvest Treatment
PROCESSING
STORAGE Conditions, temperature, time
OTHER FACTORS Diseases, deterioration, etc.
P Properties
APPEARANCE e. g., colour
TASTE e. g., palatability
PHARMACOLOGICAL EFFECTS
PHYSICAL PROPERTIES e. g., texture, acidity, pH
OTHER FACTORS
C
Chemical Compounds
UNITS
QUANTITY
Analysis
FREE: COMBINED: TOTAL
QUALITATIVE SEMI QUANTITATIVE
WHOLE: PORTION of plant part
WET: DRY: CONTROLLED
Notes
Methods of analysis
Special points
MF/

FIGURE 1. Headings for indexing entries.

In the collection, there are four main aspects: botanical, chemical, functional (e.g., palatability), and miscellaneous (e.g., cultivation). Each of these could be indexed by its own classification. For the functional and miscellaneous aspects any classification would be arbitrary, but the number of headings is small enough for this to be unimportant. There would be no need to develop botanical and chemical classifications. Well-established classifications exist. If a classified system were to be used, however, the freedom to make generic searches by any criterion, which was a main requirement of the collection, would be lost, and it was concluded that an alphabetical, classified system would not be suitable.

Alphabetical systems not being suitable, the entries must be encoded in some way. This would facilitate machine handling which the size of the collection also indicated might be desirable.

In choosing a suitable notation[3] in which to express the entries, the need for showing implicit relationships was kept well in mind. In entries made in plain language, the relationships between the words of the entry are expressed partly by special words of relation, partly by the order in which they are recorded, and partly by inflexion (i.e., special modifications) of the words used. If the "words" are reduced to single symbols, the last of these methods becomes synonymous with the use of a special "word" of relation. It should be possible to construct an artificial language, or notation, in which synonyms are rigorously excluded and in which the redundancy is reduced to a controlled amount. A separate sequence of symbols, or "word," will be required for each concept to be expressed and for each relationship between the concepts. The idea of expressing relationships in this way has been proposed by Farradane (o), who employs special symbols for the purpose, and it is also included in the colon classification (p). If every "word" is to be expressed by a single digit, an impracticably large number of characters will be required. "Words" consisting of two or more characters must be made so as to reduce the number of characters to a usable level. Reduction in number of characters has to be paid for by the complication that the meaning of a character has become dependent on its context. To reduce this complication to a minimum, the number of characters used should be as large as possible. The characters used should be easily distinguishable and reproducible and be adaptable to manuscript writing. The range of characters found on a typewriter keyboard meets these requirements. Customarily these comprise upper and lower case alphabets, one range of numerals, a set of punctuation marks, and some special characters. A more suitable modification for chemical use is a set comprising upper and lower case

[3] The term "notation" is used here to describe the system in which information is recorded in the collection, and the term "code" is reserved for the pattern of holes or spots in which the symbols of the notation are represented for machine handling.

alphabets, numerals, subscript numerals, and punctuation marks. If the typewriter is fitted with accents, accented characters increase the range available, although these might be considered as double characters since two key movements are required to reproduce them.

The subject to be indexed would be described in this notation, element by element, by taking the elements in a standard order. The resulting cipher would be a unique representation of the subject and the same subject would always be represented by the same cipher. Common sequences in the two ciphers would represent common elements and relationships. However, the position of these sequences in the ciphers will in general not be the same, owing to the "grammatical rules." The sequences will also owe their individuality to the order of the characters of which they are composed. In consequence, any system used for retrieval must be able to recognise that, for example "bad" is not equivalent to "dab."

The problem of depicting chemical compounds so that they might be indexed properly and uniquely has been under study by a number of workers during the last decade. In 1949 the International Union of Pure and Applied Chemistry invited submission of codes for chemical compounds which satisfied their requirements (z). Of the systems submitted in response to this invitation, the Dyson system (aa) has been selected after extensive testing as the Proposed International System (bb). This notation uses a large number of characters (162 in all) and in general satisfies the desiderata set out above as desirable for an indexing notation. Some of these characters, (e.g., overlined characters and fractional subscripts) could perhaps be dispensed with for normal indexing and if this is done, the notation can be expressed in approximately 107 characters.

No similar notation has been developed for plants. These have been traditionally classified in a linear order according to their main features. There are a number of anomalies in the order, and constant changes are being made in an effort to minimise these. For the flowering plants, two classifications (f, g) have found general acceptance. Another (h) based on a somewhat different starting point has been proposed more recently. Sporne (l, m) has suggested an alternative system based on the probable evolution of the plants. Two proposals have been made to describe plants by a fixed serial number (j, k). For a study of the relationship between chemical composition and the taxonomic characters of plants, it would be necessary to express these characters in a notation of the type proposed above. The outline of such a notation has been developed and this will be developed further when the problems of machine handling have been brought nearer solution.

A number of the indexing and classification systems which have been pro posed as solutions to the problem of documenting large systems for information retrieval were examined to see whether they could be employed in this collection. All the coordinate systems, for example, are unsuitable since their principle of operation is conjunction of a number of headings of equal rank. In so doing, all order of the elements is destroyed.

Punched card machines of the conventional type also are not suitable for use with such notations for two reasons. First, they can be adapted to read more than 35–40 symbols only at considerable cost. Secondly, they normally read the cards broadside on, and it is necessary to specify the column in which a given code appears or to search the cards column by column. This repeated operation seriously reduces the rate of searching. What is required is a machine which reads each code in turn and selects only those items which contain the desired sequence. Such a machine was made for the purpose of handling Dyson's chemical notation (n) and this was demonstrated by IBM in 1950. Unfortunately this prototype has not been developed further.

High operating speeds and great flexibility in operation are the characters which distinguish electronic computers. A computer is able to handle any type of notation and can be set to search for a complete specification in one pass. In addition, the calculating facilities built in enable a computer to work to alternative specifications in a way in which no other system can. Disadvantages of computers for information retrieval are their great cost and complexity and the large amount of detailed programming necessary before inserting or retrieving information. Also, in many machines input and output speeds are low in comparison with those of the calculating units. The basic needs of an information system are for a large store of data on which relatively little work will be done and for an output speed comparable with the rate at which the data searching is done, whereas the computer is most efficient when performing a number of sequential operations on the same data. A collection of the size and nature described above was likely to exceed the storage capacity of any computers existing when the survey was begun.

Microfilm rapid selectors, that is, electronic selectors in which the selection media are in microfilm form, have the advantage that they can produce the information corresponding to the search specification rather than the reference numbers of the documents which contain it. The film media are small and comparatively cheap to make and to store and are readily replicated. It seemed that further investigation of their possibilities was merited. A number of systems have been described. The Shaw rapid selector (u) redesigned and tested at the U.S. Department of Agriculture Library (t) was one of the earliest. The photoscopic information storage system (q) makes use of computer type circuits to analyse the information in the system and has the greatest density of information in its storage. It is, however, difficult as yet to amend information once it has been recorded. The Minicard system (r, s) has a large coding area, and special attention has been given to reproduction of copies, addition of codes to later prints and to the inclusion of the maximum amount of information in clear. The system also includes the fullest range of handling machines and represents the most comprehensive attempt to provide for the needs of an information system. Unfortunately it appears (s) that the codes are not to be read from end to end of the card on each pass but row by row as is done in the case of IBM cards. Another selector is in course of development at Western Reserve University (v). Unfortunately, none of these systems is currently available in the United Kingdom, and it was impracticable to base a system on any of them. There was, however, one machine of this type, the Filmorex (w, x), which is of moderate price and is currently available. This system was thought worthy of further investigation. It was accordingly chosen for trial.

One of the problems which arises whenever non-redundant codes or notations are used is that of detecting and eliminating errors. If the data are reproduced mechanically once they have been recorded and checked, errors in transcription are minimised. This can be suitably done by recording the information on punched paper tape. The Flexowriter automatic typewriter (y), which in its simplest form is an electric typewriter with a tape-punch and tape-reader attached to it, enables this to be done. This model can punch selected portions of the information onto the paper tape as it is typed. The resulting tape can afterwards be used to operate the typewriter. The codes can be punched in a strip along one edge of a card, by a modified version of the punch, and these cards used instead of the tape to operate the reader. There is also a more complex model of the machine (the Programatic). In this model, codes can be punched in the tape which will cause the machine to switch the punch and/or the typewriter on and off, enabling extracts to be made of predetermined parts of the information recorded. Another model, intended primarily for personalised letter writing and similar uses, was provided with two readers and could be switched from one reader to the other by codes in the tapes. It was thought that a combination of these two features, the two inputs and the ability to control the operation of the machine by codes in the tape, would result in a very powerful and flexible machine. It was already standard for a tape in the reader to be used as a "programme tape" to instruct the machine to move to the position required, for the next fill-in on a form for example, and whether to punch that block of information into the output tape. The second reader and the "reader switch" facility would enable a "programme tape" to be used also for determining what action should be taken on each block of information bounded by two "reader switch" codes without this action being predetermined before the information was first recorded. Not only can this "programme tape" determine whether the information is to be typed, punched, or ignored, but even the arrangement of the blocks of information on the page can be changed. For example, insertion of carriage return codes from the programme would arrange blocks of information one below the other, whereas previously they had followed one another on the same line.

Insertion of the currect notation would be ensured by preparing and checking in advance a master set of unit cards into which was "strip-punched"[4] both the plain language and the notational equivalent of all the subject headings to be used. To make an entry, the prepunched card from the master set corresponding to the correct subject heading would be chosen and read in the first reader of the Flexowriter, which would be controlled in Duplex working by a programme tape in the second reader. The cards from the master set would be refiled immediately after use so as to be available for re-use when required. This procedure would ensure: (1) that only approved terms were used since unit cards are added to the master set only for approved subject headings; (2) that the plain language entry would be accompanied by the correct notational equivalent. If therefore a typescript and a record tape were made simultaneously, and the typescript were proofread to ensure that the correct plain language headings had been entered, the notation in the record tape must be correct. At a later stage, after any necessary further editing, a run of the record tape through the machine under control of an appropriate programme tape would result in a tape containing only the code entries of the notation. There was evidence that such a system could achieve good reliability.

A machine was made to this specification, but on trial it was found to have minor shortcomings which diminished its usefulness. For example it was found that in "non-print" the machine would respond only to a "print restore" code, with the result that it was impossible to type an extract of a record tape under the control of a programme tape since the "reader switch" codes (which switched the input from one reader to the other) were ignored. Another difficulty was that the "reader switch" codes were punched from both readers so that the number in the output tape was doubled with each pass through the Flexowriter under the control of a "programme tape" in the other reader. This made it impossible to use a "programme tape" to control the machine unless it was known beforehand how many times the information had already been through the machine. These and other minor points have been rectified, and a trial is in progress on the lines indicated.

One other modification to the standard machine deserves mention. The

[4] The term "strip-punched" is used to distinguish this type of card (which has the punched tape code punched along one side) from the Keysort type "edge-punched" card and the IBM type of "field-punched" card.

standard codes for the paper tape are 5-unit, 6-unit, and 8-unit. The 5-unit code does not contain enough combinations. In the 6-unit tape, the upper and lower case characters differ only in that the former are preceded (at any distance back along the tape) by an upper case code, while a lower case code precedes the latter. In the standard Flexowriter 8-unit code, which was designed for ease of conversion of information on tape into IBM punched cards, the 6-unit code is used with an added "parity" punch in the 5th channel so that all the codes have an odd number of holes: the 8th channel is used only for the "carriage return" code. These codes were modified so that there should be a difference between upper and lower case characters in the combination being read by the reader. A punch in the 8th channel was added to all upper case characters, while this channel was blank for the lower case. This gave an 8-unit code, one channel of which was used only for parity check. The parity channel can be omitted later and a 7-unit code results. The number of non-zero combinations (127) is enough to provide a separate code for each character on the Flexowriter keyboard and one for each control signal. If the control codes were eliminated, there is accommodation for a further alphabet which could be represented in typescript by accented letters.

The products of the above procedure are: (1) an output tape in which is punched the notational equivalent of all the indexing entries to be made for a particular paper, together with (2) a typescript on which is typed the plain language of all these entries. These are to be passed to the Filmorex for use as follows: the tape to produce the perforated mask from which the code pattern is photographed; and the typescript, together with a suitable abstract of the paper, to be photographed as the pictorial portion of the Filmorex "fiche."

A special conversion unit is required to link the Flexowriter and the Filmorex. This unit reads the codes in the Flexowriter output tape, recodes them as appropriate, and punches the new codes into the Filmorex perforated mask. By varying the connections in this unit, the coupling can be made flexible. The principle of operation of the Filmorex selector, of passing the cards in turn through a beam of light, in which is placed also a search specification card bearing the inverse of the pattern sought and of using the momentary "black out" of all light which occurs when a wanted card is read to operate the selection shutter by means of a photocell, imposes a limitation on the codes which can be used. Each pattern (in the standard Filmorex, one line of the coding area) read by a single photocell must have the same number of black spots (and of white spaces). With this limitation, the coding area, 30 units wide, can be divided into 6 fields each 5 units wide allowing a 6-digit number to be represented (2 punches out of 5 give 10 characters). The possible vocabulary size is 10^6 wrods. Alternatively a larger vocabulary (3.2×10^6 words) can be used by dividing the line into 5 fields of 6 units each (punching 3 out of 6 gives 20 characters and 20^5 possible “words”). Since the output of the Flexowriter is 7-unit alphanumerical, it was thought preferable to divide the line into 4 digits of 7 units each, allowing 35 characters to be represented (punching 3 out of 7). The vocabulary is reduced somewhat below the maximum (to approximately 1.5×10^6) but greater flexibility is achieved. The 35 characters chosen are the numerals plus an alphabet, omitting I as likely to conflict with 1.

As the equipment stands, the system is simple and flexible. The codes are read in order on successive lines of the fiche, and the extensive presorting which the small, cheap fiche makes economic, speeds up the search by making it unnecessary in the majority of cases to search more than a small fraction of the file. The present selector has 5 reading heads and can be set to select various logical combinations of 5 code lines at one pass. It cannot distinguish the order in which these code lines occur.

If, however, the reading mechanism of the Filmorex were altered by the addition of further photocells, it would be possible to remove the restriction mentioned above on the code combinations which can be used, and the full theoretical total of 127 non-zero combinations possible for a 7-unit code could be used to accommodate 127 different characters or control signals. The versatility of the selector could be further increased by adding more logical circuits. By using the two spare units to enable the number of lines to be counted, this logical circuitry could distinguish the order of codes.

In order to allow a trial of the collection to be begun, it was decided to accept the present limitations of the Flexowriter and Filmorex and to try them in combination before attempting any further modifications, such as those outlined above. These can be worked on as the above trial is in progress. For work to commence, a botanical and chemical notation must be worked out which does not need more than 35 characters.

For the botanical entries, a system has been worked out. Zero was reserved for generality, and it was found that the whole of the plant kingdom could be accommodated (Table 1). For the flowering plants a more detailed classification has been made by using Willis's system (i) as a basis. Nine alphanumeric digits are used. Of these the first three designate the family (Table 2), the next two the genus, and the remaining four the species and variety. So far some 5,000 species of flowering plants have been satisfactorily coded. For ease in recognition, a space is left after the third digit (the family) and a decimal point is inserted after the fifth (the genus), e.g., 365 72.11ØØ. (Zeros are barred to distinguish them from the letter O.)

For the chemical compounds, several notations have been proposed which can be expressed within the limits of 35 characters. The Wiswesser system (cc, dd, ee) was originally adapted to the slightly greater range of characters which a punched card machine can handle but has been developed with more characters into a notation such as was envisaged by IUPAC (z) and is designed for correlation and searching procedures. The Chemical-Biological Coordination Center developed a code (ff) for use with its work which is of a completely different character. In this the various component groupings are enumerated and no attempt is made to designate the complete compound with a unique cipher. The Centre National de la Recherche Scientifique has developed a

TABLE 1. Plant classification: major groups

CRYPTOGAMIA	
Thallophyta	
Bacteria	A
Myxomycetes	B
Algae	
Chlorophyceae	C-D
Xanthophyceae	E
Bacilliarophyceae	F
Euglenineae	G
Phaeophyceae	H
Rhodophyceae	I
Cyanophyceae	J
Fungi	
Phycomycetes	
Oomycetes	K
Zygomycetes	L
Ascomycetes	
Endomycetales	M-N
(Yeasts as such)	N
Plectomycetes	O
Discomycetes	P
Pyrenomycetes	Q
Basidiomycetes	
Ustilaginales	R
Uredinales	S
Hymenomycetes	T
Gasteromycetes	U
Fungi imperfecti	V
Lichenes	
Ascolichenes	W
Bryophyta	
Hepaticae	X
Muscineae	Y
Pteridophyta	Z
PHANEROGAMIA	
Spermaphyta	
Gymnospermae	110–140
Angiospermae	
Monocotyledonae	170–200
Dicotyledonae	
Archichlamydeae	300–700
Sympetalae	860–900

TABLE 2. *Families of flowering plants and ferns (Willis)*

Pteridophyta	
Cyatheaceae	Z12
Equisetaceae	Z31
Gleicheniaceae	Z16
Hymenophyllaceae	Z11
Isoetes	Z99
Ligulatae	Z50
Lycopodiaceae	Z41
Marattiaceae	Z23
Marsiliaceae	Z21
Matoniaceae	Z15
Ophioglossaceae	Z24
Osmundaceae	Z18
Parkeriaceae	Z14
Polypodiaceae	Z13
Psilotaceae	Z71
Salviniaceae	Z22
Schizaeaceae	Z17
Gymnospermae	
Cycadaceae	111
Ginkgoaceae	121
Gnetaceae	141
Pinaceae	132
Taxaceae	131
Monocotyledons	
Alismaceae	185
Amaryllidaceae	275
Aponogetonaceae	183
Araceae	241
Bromeliaceae	259
Burmanniaceae	291
Butomaceae	186
Cannaceae	283
Centrolepidaceae	253
Commelinaceae	261
Cyanastraceae	263
Cyclanthaceae	231
Cyperaceae	212
Dioscoreaceae	278
Eriocaulaceae	256
Flagellariaceae	251
Gramineae	211
Haemodoraceae	274
Hydrocharitaceae	187
Iridaceae	279
Juncaceae	271
Lemnaceae	242
Liliaceae	273
Marantaceae	284
Mayacaceae	254
Musaceae	281
Najadaceae	182
Orchidaceae	292
Palmae	221
Pandanaceae	172
Philydraceae	264
Pontederiaceae	262
Potamogetonaceae	181
Rapateaceae	258
Restionaceae	252
Scheuchzeriaceae	184
Sparganiaceae	173
Stemonaceae	272
Taccaceae	277
Thurniaceae	257
Triuridaceae	191
Typhaceae	171
Velloziaceae	276
Xyridaceae	255
Zingiberaceae	282
Dicotyledons	
Acanthaceae	951
Aceraceae	657
Achariaceae	741
Achatocarpaceae	487
Actinidiaceae	712
Adoxaceae	974
Aextoxicaceae	664
Aizoaceae	488
Akaniaceae	663
Alangiaceae	778
Amarantaceae	482
Anacardiaceae	645
Ancistrocladaceae	746
Anonaceae	524
Apocynaceae	915
Aquifoliaceae	649
Araliaceae	791
Aristolochiaceae	461
Asclepiadaceae	916
Balanophoraceae	458
Balanopsidaceae	361
Balsaminaceae	667
Basellaceae	492
Batidaceae	391
Begoniaceae	745
Berberidaceae	518
Betulaceae	421
Bignoniaceae	934

Bixaceae	729
Bombacaceae	686
Boraginaceae	924
Bretschneideraceae	659
Brunelliaceae	559
Bruniaceae	563
Brunoniaceae	993
Burseraceae	622
Buxaceae	641
Byblidaceae	558
Cactaceae	751
Callitrichaceae	639
Calycanthaceae	522
Calyceraceae	995
Campanulaceae	991
Canellaceae	732
Capparidaceae	532
Caprifoliaceae	973
Caricaceae	742
Caryocaraceae	717
Caryophyllaceae	494
Casuarinaceae	311
Celastraceae	651
Cephalotaceae	555
Ceratophyllaceae	512
Cercidiphyllaceae	515
Chenopodiaceae	481
Chlaenaceae	682
Chloranthaceae	323
Cistaceae	728
Clethraceae	861
Cneoraceae	618
Cochlospermaceae	731
Columelliaceae	939
Combretaceae	779
Compositae	996
Connaraceae	574
Convolvulaceae	921
Coriariaceae	643
Cornaceae	793
Corynocarpaceae	648
Crassulaceae	554
Crossosomataceae	572
Cruciferae	533
Crypteroniaceae	772
Cucurbitaceae	981
Cunoniaceae	561
Cynocrambaceae	484
Cynomoriaceae	787
Cyrillaceae	646
Daphniphyllaceae	638
Datiscaceae	744
Desfontainiaceae	913
Diapensiaceae	866
Dichapetalaceae	636
Diclidantheraceae	896
Didieraceae	661
Dilleniaceae	711
Dipsacaceae	976
Dipterocarpaceae	723
Droseraceae	543
Dysphaniaceae	493
Ebenaceae	892
Elaeagnaceae	765
Elaeocarpaceae	681
Elatinaceae	724
Empetraceae	642
Epacridaceae	865
Ericaceae	864
Erythroxylaceae	616
Eucommiaceae	565
Eucryphiaceae	713
Euphorbiaceae	637
Eupomatiaceae	525
Fagaceae	422
Flacourtiaceae	735
Fouquieraceae	727
Frankeniaceae	725
Garryaceae	341
Geissolomataceae	761
Gentianaceae	914
Geraniaceae	611
Gesneriaceae	938
Globulariaceae	942
Gomortegaceae	527
Gonystilaceae	683
Goodeniaceae	992
Grubbiaceae	454
Guttiferae	722
Gyrostemonaceae	486
Haloragidaceae	785
Hamamelidaceae	564
Hernandiaceae	52x
Heteropyxidaceae	766
Himantandraceae	513
Hippocastanaceae	658
Hippocrateaceae	652
Hippuridaceae	786

Hoplestigmataceae	734
Humiriaceae	615
Hydnoraceae	463
Hydrocaryaceae	783
Hydrophyllaceae	923
Hydrostachyaceae	553
Icacinaceae	656
Juglandaceae	381
Julianaceae	411
Labiatae	926
Lacistemaceae	324
Lactoridaceae	523
Lardizabalaceae	517
Lauraceae	529
Lecythidaceae	775
Leguminosae	575
Leitneriaceae	371
Lennoaceae	863
Lentibulariaceae	941
Limnanthaceae	644
Linaceae	614
Lissocarpaceae	895
Loasaceae	743
Loganiaceae	912
Loranthaceae	457
Lythraceae	771
Magnoliaceae	521
Malesherbiaceae	738
Malpighiaceae	631
Malvaceae	685
Marcgraviaceae	718
Martyniaceae	936
Medusagynaceae	716
Melastomataceae	782
Meliaceae	623
Melianthaceae	666
Menispermaceae	518
Monimiaceae	528
Moraceae	432
Moringaceae	536
Myoporaceae	953
Myricaceae	351
Myristicaceae	526
Myrothamnaceae	562
Myrsinaceae	872
Myrtaceae	781
Myzodendraceae	451
Nepenthaceae	542
Nolanaceae	931
Nyctaginaceae	483
Nymphaeaceae	511
Nyssaceae	777
Ochnaceae	714
Octoknemataceae	456
Olacaceae	455
Oleaceae	911
Oliniaceae	763
Onagraceae	784
Opiliaceae	453
Orobanchaceae	937
Oxalidaceae	612
Pandaceae	581
Papaveraceae	531
Passifloraceae	739
Pedaliaceae	935
Penaeaceae	762
Pentaphylacaceae	647
Phrymaceae	954
Phytolaccaceae	485
Piperaceae	322
Pittosporaceae	557
Plantaginaceae	961
Platanaceae	571
Plumbaginaceae	881
Podostemaceae	551
Polemoniaceae	922
Polygalaceae	635
Polygonaceae	471
Portulacaceae	491
Primulaceae	873
Proteaceae	441
Punicaceae	774
Pyrolaceae	862
Quiinaceae	719
Rafflesiaceae	462
Ranunculaceae	516
Resedaceae	535
Rhamnaceae	671
Rhizophoraceae	776
Rosaceae	573
Rubiaceae	971
Rutaceae	619
Sabiaceae	665
Salicaceae	331
Salvadoraceae	653
Santalaceae	452
Sapindaceae	662
Sapotaceae	891

system on somewhat the same lines as the CBCC for use in its Filmorex installation (gg) and it is proposed to adopt this for the first trial run.

Sarraceniaceae	541	Theophrastaceae	871
Saururaceae	321	Thymelaeaceae	764
Saxifragaceae	556	Tiliaceae	684
Scrophulariaceae	933	Tovariaceae	534
Scytopetalaceae	688	Tremandraceae	634
Simarubaceae	621	Trigoniaceae	632
Solanaceae	932	Tristichaceae	552
Sonneratiaceae	773	Trochodendraceae	514
Stachyuraceae	736	Tropaeolaceae	613
Stackhousiaceae	654	Turneraceae	737
Staphyleaceae	655	Ulmaceae	431
Sterculiaceae	687	Umbelliferae	792
Strasburgeriaceae	715	Urticaceae	433
Stylidiaceae	994	Valerianaceae	975
Styracaceae	894	Verbenaceae	925
Symplocaceae	893	Violaceae	733
Tamaricaceae	726	Vitaceae	672
Theaceae	721	Vochysiaceae	633
		Zygophyllaceae	617

Other properties, e.g., palatability, medicinal effects, texture, conditions of growth, susceptibility to diseases, are of importance in determining the economic value and use which can be made of plants. These properties are difficult in many cases to describe on a numerical or other linear scale. The number of possible headings for each is small, however, and coding in a maximum of two digits is possible in a number of ways, within the range of 35 characters.

At the time when this paper was proposed, it was thought that all the above experimental work could be reported on. Owing to delays in delivery of equipment and, in particular, to a serious accident which befell the author, it has not proved possible to include the results. The present status of the work is: the Flexowriter has been tried on all the procedures, and, subject to the modifications outlined, has proved itself satisfactory; the detailed design of the conversion unit has been completed and construction is due to begin. The delivery of the Filmorex equipment is expected in the near future. By autumn 1958 some results should be available and these will be reported in due course.

SUMMARY

Some serious limitations of existing methods of indexing and cataloguing scientific information became apparent when the possibility was being explored of setting up a large detailed system which would answer enquiries on the pub

lished information on the chemical compounds found in plants. It was estimated that this system would need to contain of the order of 10^7 items, each of which might be sought from any one of four aspects, namely chemical, botanical, functional, and miscellaneous. By functional is meant such characters as palatability, pharmaceutical effects, and toxicity while the miscellaneous aspect includes such factors as cultivation and place of growth. The problem was complicated by a need for retrieval when given partial specifications, for example plants containing chemical compounds which have certain groupings in common. Such a system cannot be handled by any of the existing classical methods.

The solution proposed, which is applicable to any system of large size which has to be indexed in detail, is to express each of the factors in a notation in which a linear series of symbols expresses the factor element by element and in which the relationship between the elements is expressed partly by the order of these symbols and partly by special symbols of relationship. In the example mentioned above, notations for expressing the chemical aspect have been proposed, the botanical aspect has been extensively classified by taxonomic characters although no notation of this type has been formulated, while the functional aspect is not classified.

Use of such a notation on the scale envisaged presents problems of machine design. Although some machines work satisfactorily on a binary system, a range of symbols expressed in binary form is not convenient for manual handling (e.g., compilation of codes and entry of the information into the system). For this, use of the maximum number of symbols is desirable. A compromise has been adopted with a range of symbols as large as can be accommodated on a typewriter keyboard, that is to say, two alphabets, two ranges of numerals, and a full set of punctuation marks, which can be converted by a modified model of a tape punching typewriter into a seven-bit code.

Having recorded the information in this way and having checked it for accuracy, further handling of the information can be done by machine and further checking should be unnecessary. Various devices can be used to facilitate this initial checking.

Selection again presents a problem. The standard punched card machine, which immediately comes to mind as a possible way of mechanical selection, is unsuited to seven-bit codes and even more unsuited to notations of the type proposed in which the position of the group to be searched for cannot be specified. On the other hand, electronic computers, which are ideal for handling binary information, have memories which are several orders too low and have the additional disadvantage of being unnecessarily expensive.

The selection can be done by a rapid selector type of equipment modified so as to accept punched tape (the output from a tape punching typewriter) as its input. In this machine the indexing entries recorded in binary code are scanned term by term so that the position of the group for which search is being made is immaterial.

There are three points of novelty in the process. First is the development of suitable notations and their adaptation to machine limitations and use. Second is the modification to the tape punching typewriter to give it the necessary versatility and flexibility. Third is the adaptation of the Filmorex to accept codes of this type and the design and construction of the converters needed to transfer the information from one machine to the next.

REFERENCES

(a) *Survey of Agricultural, Forestry and Fishery Products in the United Kingdom.* Development Commission, London, 1953.

(b) KLEIN, G. *Handbuch der Pflanzenanalyse.* Springer, Wien, 1931–2.

(c) WEHMER, C. *Die Pflanzenstoffe.* Verlag Fischer, Jena, 1929–30.

(d) WALL, E.M., FENSKE, C.S., WILLAMAN, J.J., CORRELL, D.S., SCHUBERT,B. G. and GENTRY, H.S. Steroidal sapogenins XXVI. Supplementary table of data for steroidal sapogenins XXV. September 1955. U.S. Dept. Agr., Agric. Research Service, ARS-73–4.

(e) URQUHART, D.J. *Unanswered Questions No. 5.* Dept. Scientific Industrial Research, London, October 1951, pp. 1–3.

(f) BENTHAM, G. and HOOKER, J.D. *Genera Plantarum,* London, 1862.

(g) ENGLER, A., and PRANDTL. *Die Natürlichen Pflanzenfamilien,* Leipzig.

(h) HUTCHINSON, J. *Families of Flowering Plants:* Vol. I, *Dicotyledons;* Vol. II, *Monocotyledons.* Macmillan, London, 1926/34.

(i) WILLIS, J.C. *Flowering Plants and Ferns,* 6th revised edition. Cambridge University Press, 1951.

(j) MULLINS, L.J., and NICKERSON, W.J. A proposal for serial number indentification of biological species. *Chronica botan., 12,* 4 (1951).

(k) GOULD, S.W. Permanent numbers to supplement the binomial system of nomenclature. *Am. Scientist, 42,* 269–74 (1954).

(l) SPORNE, K.R. Statistics and the evolution of dicotyledons, *Evolution, 8* [1], 55–64 (1954).

(m) SPORNE, K.R. The phylogenetic classification of the Angiosperms. *Biol. Revs., 31,* 1–29 (1956).

(n) DYSON, G.M. Studies in chemical documentation III. Mechanized documentation. *Chem. & Ind., 1954* (April 17), 400–9.

(o) FARRADANE, J.E.L. A scientific theory of classification and indexing and its practical applications. *J. Document. 6,* 83–99 (1950); A scientific theory of classification and indexing: further considerations. *J. Document., 8*(2), 73–82 (1952).

(p) RANGANATHAN, S.R. *Annals of Library Science.*

(q) KING, G.W. A new approach to information storage. *Control Eng.,* August 1955. KING, G.W., BROWN, G.W., and RIDENOUR, L.N. Photographic techniques for information storage. *Proc. I.R.E., 41*(10), 421–8 (1953). ANON. Photoscopic information storage. International Telemeter Corporation, Los Angeles, Calif., Publication R-77, March 15, 1955.

(r) TYLER, A.W., MYERS, W.L., and KUIPERS, J.W. The application of the Kodak Minicard system to problems of documentation. *Am. Document., 6*(1), 18–30 (1955). KUIPERS, J.W., TYLER, A.W., and MYERS, W.L. A Minicard system for documentary information. *Am. Document., 8*(4), 246–68 (1957).

(s) Minicard demonstration. *Am. Document., 6*(4), 258–9 (1955).

(t) Report for the microfilm rapid selector. Eng. Research Assoc., Inc., No. PB 97313. U.S. Dept. Commerce, 1949.

(u) SHAW, R.R. The rapid selector. *J. Documentation, 5,* 164–71 (1949).

(v) The Western Reserve Searching Selector. *Am. Document, 8*(3), 237–8 (1957).

(w) SAMAIN, J. Progres du classement et de la selection mechanique des documents. 17 Conf. FID Berne, August 1947, pp. 22–26.

(x) SAMAIN, J. The organization of documentation by the Filmorex technique. Filmorex, Paris, 1956.

(y) BROWN, R.HUNT, Editor, *Office Automation.* Automation Consultants, Inc., New York, 1955, pp. 51–7.

(z) Codes invited. *Chem. Eng. News, 27,* 2998 (1949).

(aa) DYSON, G.M. A New Notation and Enumeration System for Organic Compounds, 2nd edition. Longmans Green, London, 1949.

(bb) DYSON, G.M. Private communication.

(cc) WISWESSER, W.J. Simplified chemical coding for automatic sorting and printing machinery. Willson Products Inc., Reading, Pa., 1951.

(dd) WISWESSER, W.J. The Wiswesser line formula notation. *Chem. Eng. News, 30,* 3525–6 (1952).

(ee) WISWESSER, W.J. *A Line Formula Chemical Notation.* Crowell Co., New York, 1954.

(ff) CHEMICAL-BIOLOGICAL COORDINATION CENTER. A method for coding chemicals for correlation and classification. National Research Council, Washington, D.C. 1950.

(gg) SAMAIN, J. Personal communication.

TABLEDEX: A New Coordinate Indexing Method for Bound Book Form Bibliographies

ROBERT S.LEDLEY

1. INTRODUCTION

The conventional bound book form of bibliography is an important class of publication, widely used in the sciences, arts, industry, business, and government. The publication of such a book is routine and has the additional advantages of convenience and familiarity. However, there is a serious disadvantage to a bibliography of this form: the indexing system commonly used is inadequate for *coordinate* retrieval. That is, one cannot easily look up all articles[1] in the bibliography associated with several (three or more) descriptive words. In many cases this severely limits the utilization and value of the bibliography as a reference source.

The purpose of this paper is to describe a new method for indexing a conventional bound book form of bibliography, called Tabledex, which will enable easy and rapid complete coordinate retrieval. That is, by means of this new indexing method, all articles listed in the bibliography and associated with any number of descriptive words can be found easily and rapidly.

For example, in a bibliography of current research reports suppose it is desired to look up all articles on “laboratory aids to the differential diagnosis of hematological diseases.” With a conventional index one might start by looking under “hematological diseases” in the hope of finding “diagnosis”; if this failed, one might then look under “diagnosis” for “hematological diseases,” and most likely the “differential” and “laboratory aids” aspect will have been completely lost. On the other hand, if the bibliography has a coordinate index of the kind described in this report, then it would be an easy, straightforward

ROBERT S.LEDLEY Data Processing Division National Bureau of Standards, Washington, D.C.

[1] We are using the word “article” genetically to mean all titles and references that may be listed in a bibliography.

matter to look up all articles on "laboratory aids to the differential diagnosis of hematological diseases." One would automatically find all articles associated with *all* of the words: *laboratory, aids, differential, diagnosis, hematological,* and *disease*.

Background. The coordinate retrieval concept has long been known in the field of indexing, and several methods for manually implementing coordinate retrieval of information are well known. All these require a certain amount of manipulation of cards or pages, and are unfortunately not appropriate for use as an index in a bound book. Coordinate retrieval has been accomplished by using marginal punched cards, where each card represents an article, and each marginal position a word; on each card the words that characterize the corresponding article are recorded by notching the respective marginal position; coordinate retrieval is accomplished by separating cards from such a deck by means of the notches corresponding to a set of coordinated words. On the other hand, another technique, often called the Peek-a-boo method, uses a card to represent a word, and the articles are represented by respective *x-y* coordinate positions on the card; on each card, the articles that are characterized by the corresponding word are recorded by punching a hole at the respective *x-y* coordinates; coordinate retrieval is accomplished by placing together the cards from the deck that correspond to a set of coordinated words and then observing which *x-y* coordinate positions are punched on every selected card. Finally, in the Uniterm split book method, on each half page appears one or more words, and numbers that represent articles characterized by each word; the half pages corresponding to a set of coordinated words are scanned for common numbers. The special nature of this book departs enough from the usual publishing format to cause extra problems for the publisher. The method has the additional disadvantages for our purposes that when making a single coordination with more than two words, one must refer to several pages, and must also copy down on scrap paper the article numbers that appear on the intermediate steps.[2]

Advantages of Tabledex. There are two fundamental advantages of this new method of indexing. The first is that it maintains the convenience of the conventional book form together with easy and rapid coordinate retrieval.[3] *No* manipulation of cards is necessary; *no* artifacts, such as holes in the paper, are needed; it is not necessary to compare two different pages, nor copy down intermediate numbers. All the work is done on a single page of the book. Of

[2] Of course, the Uniterm method is well suited for other purposes, particularly because it is essentially open-ended.

[3] Some advantages that would result from non-manipulative coordinate indexes have been discussed in the literature, cf. Charles Bernier, Correlative Indexes I. Alphabetical Correlative Indexes. *American Documentation, 7*[4], (1956).

course, a bound book form of bibliography implies a *closed* indexing method, because additional articles or words cannot be included in the bibliography once it is published.

Use of computers. The second fundamental advantage is that, with the new method, making a bibliography can be highly mechanized, and the actual work of the bibliographer is reduced to a minimum. That is, a computer (such as a digital electronic computer) can, within a few hours, automatically organize, compile, and print the entire bibliography including the Tabledex coordinate index—the computer will do everything described in section 3 of this paper even to making up the page formats. The bibliographer need only choose the words that are associated with each article included in the bibliography, which, of course, he must do no matter what system is used. Thus the work of the bibliographer is reduced solely to making human decisions—the computer will do the rest. The process might involve three steps: (1) The associations of words for articles is accomplished by the bibliographer. (2) These associations, after being transcribed onto an appropriate medium, are read into the computer. (3) Within a few hours the pages of the book bibliography are printed out by the computer, and then reproduced directly by the photo-offset process and bound into books.

It is difficult to estimate the number of pages such a bibliography will require. However, we might guess that if a bibliography contains 10,000 articles, 6000 words (3000 words with an average of 2 synonyms per word, see below), and if an average of 10 words is associated with an article, and if the print is similar to that of *Webster's New Collegiate Dictionary,* 1953, then there would probably be approximately 300 pages in the bibliography.

2. THE NEW INDEXING METHOD AND ITS USE

THREE PARTS

The book itself would consist of three parts. Part I is the bibliography proper, or list of articles with the author, journal, etc., and if desired some descriptive material. The article list designates each article with an underlined article number. Part II is the alphabetical list of descriptive words by means of which the retrieval is accomplished. The word list designates each word with a nonunderlined word number. Part III contains the *indexing tables*. For purposes of illustration, we have reproduced in Section 5 a short bibliography composed of twelve articles in Part I (the article list) and twenty-two words in Part II (the word list). (This illustrative bibliography is obviously incomplete and is used as an example only.)

THE TABLES

The actual work of looking up the desired articles is done by means of the tables in Part III. These contain two types of numbers, the underlined *article numbers* down the left most column, and the nonunderlined *word numbers* comprising the rows. There is *one table for each* (distinct) *word of the word list* and each table is numbered with this word number. The article numbers in the left most column are all articles associated with that word, and the other numbers in each row are other words which describe or are significant to this article.

USING THE NEW TABLEDEX INDEXING METHOD

An example of the actual method of using the tables can easily be seen by a simple example. Suppose one wants to find all articles on the application of nuclear theory, i.e., all articles each of which is associated with *all* of the following given words:

application

nuclear

theory

(1) From these given words, one first looks in the word list for the numbers designating each of them, and writes these given numbers in numerical order. (2) One then takes the smallest of these given numbers and turns to the table of that number. (3) One inspects this table, searching for a row which contains all the other given numbers. Check mark this row and each row that also has all the other given numbers. Since each of these rows is labeled with the underlined article number on the left, these are all the articles one is looking for, i.e., all articles associated with all these words. For example:

1. From the alphabetical word list one finds: application 5. 1, nuclear 4. 3, theory 4.4, which in numerical order are 4.3, 4.4, and 5.1.
2. Then one turns to Table 4.3, which is reproduced for convenience.
3. Noting that only the first two rows of Table 4.3 contain both 4.4 and 5.1 these rows are checked. Hence articles <u>2. 4</u> and <u>3. 4</u> by Seidle and by Tove are associated with all the given words.

Table 4.3

√ <u>2.4</u>	4.4	5.1	5.3	6.2
√ <u>3.4</u>	4.4	5.1		
<u>3.3</u>	4.4	5.3	6.1	
<u>2.1</u>	6.2	7.1		

FURTHER OBSERVATIONS

In Part I, the articles can be listed in any convenient order, i.e., alphabetical by author, or title, or journal. (In our illustrative bibliography they are listed by author.) The assignment of article numbers need only be sequential, so that the article designated by a given article number can be easily located. In Part II, the words are listed in alphabetical order. To maximize the probability of a person finding the word he is looking for in the word list, synonyms of the descriptive words are also included in the list; but synonyms have the same word number. In this sense, the list of words is a type of thesaurus.

In Part III, as noted above, the article numbers that label the rows of a table correspond to all the articles associated with the word of the table. Hence a quick glance at a table can present *all* articles associated with the corresponding word. The word numbers appearing within a row correspond to words associated with the article of that row. For example, observation of Table 4.3 shows that 2. 4, 3. 4, 3. 3, and 2. 1 are all the articles associated with word 4. 3. Also observe, for example, that words 4. 4, 5. 1, 5. 3, and 6. 2 are associated with article 2. 4. However, as discussed in further detail in the next section, these are *not* all the words associated with article 2. 4.

CHECKING THE ROWS

After one has turned to the table of the smallest given word number, checking the rows for the other given word numbers is often best accomplished in steps of one number at a time as follows. One begins by searching for the second smallest given word number in each successive row of the table. A check mark is placed next to those rows that contain this number. One next searches for the third smallest given word number, but the search is made *only in those rows that have been previously check-marked*. The check is crossed off when this given word number does not appear; another check is marked when it does appear. One continues this process for each of the remaining given word numbers, each time considering *only* checked rows whose checks have not been previously crossed off. When this has been accomplished for all the remaining given word numbers, *those rows that are checked, and whose checks have not been crossed off correspond to articles that are associated with all of the given words*.

For our example, after turning to Table 4.3 (see Section 5) search for the second smallest given word number, namely 4. 4; check the first three rows. Next search in these rows for the third smallest given word number, namely 5. 1; place another check next to the first two rows; cross off the check of the third row. Thus rows corresponding to articles 2. 4 and 3. 4 (by Seidle and by Tove) are the desired rows, and hence these are all articles associated with all three of the given words.

It is important to note that *within each row* the numbers are always increasing to the right, which significantly aids the search. Also note that the numbers in the first column are *increasing* downward; hence *when the first number of a row is larger than the second given word number, this row and all rows below need NOT be searched.*

Tracing paper can aid this process of making the check marks and crossing off checks. The paper is placed over the page on which the proper table is located and pushed into the page binding. The table is read through the tracing paper, and the markings are made directly on it.

3. COMPILING THE NEW INDEXING SYSTEM: TABLEDEX

THE MATRIX

A conceptual visualization that often clarifies retrieval problems is the article-word matrix. The rows of this matrix are labeled by the articles under consideration, and the columns are labeled by the descriptive words. An element of this matrix is a *unit* if the *article* corresponding to the element's *row* is associated with the *word* corresponding to the element's *column;* otherwise the element is a *zero*. Of course, in practice such a matrix is never actually

		1.1 hysteresis	1.2 mass	2.1 counting	2.2 gas	2.3 thermal	3.1 differentiation	3.2 versatility	4.1 analysis	4.2 Netherlands	4.3 nuclear	4.4 theory	5.1 application	5.2 England	5.3 instrumentation	6.1 evaluation	6.2 concept	7.1 design
1.1	Abrahams	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0
1.2	Aravindakshan	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	1
1.3	Gasstrom	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0
1.4	NBS Circular	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	1
2.1	Nicholls	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1
2.2	Pope	0	1	0	1	0	1	0	1	0	0	0	1	1	0	1	0	1
2.3	Powell	0	0	0	0	1	1	1	0	0	0	0	1	1	1	0	0	1
2.4	Seidle	0	0	1	0	0	0	0	0	1	1	1	1	0	1	0	1	0
3.1	Senior	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
3.2	Smith	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1
3.3	Stockendal	0	0	0	0	0	0	0	1	1	1	1	0	0	1	1	1	0
3.4	Tove	1	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0

FIGURE 1. The article-word association matrix. Short bibliography of articles on instrumentation (from the files of the Office of Basic Instrumentation of the National Bureau of Standards).

made, but the concept aids the visualization of many coordinate retrieval methods. Such a matrix for our illustrative bibliography is given in Fig. 1 so that the description of the system may be more easily understood. Marginal punch cards (where a card represents an article and the marginal positions represent words) are visualized as representing the rows of this matrix. Peek-a-boo cards (where a card represents a word and the holes, articles) are visualized as representing the columns of this matrix. In any event, given a word, the associated articles can be easily read from the matrix and, conversely, given an article, the associated words can be easily read from the matrix.

ARTICLE NUMBERS AND WORD NUMBERS

Although we have mentioned article and word numbers above, we have not yet discussed in detail how they are composed. Let us start with the article numbers. As was noted above, the assignment of article numbers to the articles need be only sequential, according to the article list in Part I of the bibliography, so that the article designated by a given article number can be easily located. For example, in our illustrative bibliography the article number represents the page and line on which the article is located. Hence, the article corresponding to article number 2. 3 is located on page 2, line 3 of the bibliography.

The word numbers, on the other hand, are assigned in a very specific manner, in order to reduce the tables to a *minimum* size. The part of a word number to the left of the decimal point is equal to the *number of articles* listed in the bibliography that are associated with this word; the part to the right of the decimal point is meant to distinguish between different words that are to be associated with the same *number* of articles. For example, if three different words are each associated with 5 articles, then their numbers would be 5. 1, 5. 2, and 5. 3. The fact that the word number gives the number of articles associated with the word can often be directly used to great advantage.

MAKING THE TABLES

As has been mentioned, there is one table for each word number; and each table has one row for every article associated with the word. Each row is labeled on the left with the article number. For example, consider the table for the word *nuclear,* bearing the word number 4. 3. From the matrix it is seen that this word is associated with articles 2. 1, 2. 4, 3. 3, and 3. 4 hence these article numbers label the rows of Table 4.3.

The nonunderlined numbers of a row consist of all the word numbers associated with this article *that are greater than the word number of the table,* arranged in increasing order from left to right. For example, in Table 4.3 consider the row for article 2. 4. From the matrix it can easily be seen that this article is associated with word numbers 2. 1, 4. 2, 4. 3, 4. 4, 5. 1, 5. 3, and 6. 2. However, since the word number of the table is 4. 3, then only word numbers 4.4, 5.1, 5.3, and 6.2 (which are greater than 4. 3) appear in the row. Finally, the rows of a table are rearranged so that the first column of word numbers increases from top to bottom.

The purpose of making the tables in the above manner is to reduce the tables to a minimum size. Since according to the system, the given word numbers are first arranged in numerical order, and the table corresponding to the *smallest* word number is then used for retrieval, within the table the word numbers need never be smaller than the word number of the table. Hence only word numbers larger than the word number of the table need be included in the rows of the table. In addition, tables that have many rows, i.e., tables for words associated with many articles, will be very thin; since such tables will correspond to words having large word numbers, from which it follows that the rows will contain few word numbers. (For example, see Table 6. 2.) On the other hand, tables that have many words within a row will probably correspond to small word numbers, and hence they will have few rows. (See Table 1.1.) The critical tables are those for words associated with an average number of articles. Here the search for numbers within a row is simplified by arranging the numbers in increasing order; and the necessity for searching all rows has been eliminated in many cases by making the numbers in the first column increase from top to bottom.

The process of making the tables for a particular bibliography might be difficult to accomplish by manual means, but it is easy to do with the aid of a high-speed digital electronic computer. The input to the computer need be only the list of articles with their respective associated words. The computer will automatically assign both article numbers and the word numbers, as well as form all the tables. The output from the computer will be the completed bibliography, Parts I, II, and III, printed even with the correct page formats, ready for photo-offset duplication and binding into books. All this can be accomplished by the computer within a few hours.

In Appendix I we have tried to give an indication of the mechanical feasibility and advantages of the new technique. In Appendix II we have considered an additional problem: what to do if no article exists in the bibliography that is characteristic of *all* given words. In Appendix III we have suggested an important alternative approach to making index tables: using words instead of numbers. The approach is still based on the same principles developed in this article. Appendix IV discusses the advantages and disadvantages of population vs. substantive meaning for the entries in the tables.

4. APPLICATIONS OF TABLEDEX

1. For rapidly preparing bibliographies of current periodical literature

The new indexing system would not only provide a means for highly mechanizing the production of such a publication but it would also provide a coordinate index. For example, consider the monthly "Current List of Medical Literature" compiled by the National Medical Library. Articles from thousands of journals are indexed each month in this publication. Mechanizing some of the tasks of the bibliographers might have advantages, such as further reducing time delays in compilation and publication, and eliminating human errors. In addition, a coordinate index can often enhance the value and effectiveness of such a publication as, for example, when trying to find articles on such subjects as "laboratory aids to the differential diagnosis of hematological diseases," or "incidence of post-partum hemorrhage due to retained placental fragments."

2. In handbooks and other compilations

A coordinate index of this type can often enhance the value of a handbook. For example, it might be very difficult to find quickly within a handbook of mathematical formulas and tables some particular formulas that give "continued fraction expansions of the arctan x converging rapidly for $0<x<1$," or "Fourier series expansion for reciprocals of Jacobian elliptic functions."

3. As an annual index for a current journal

Many scientific periodicals, such as the *Chemical Abstracts* and *The Physical Review* have annual, semiannual, or quarterly cumulative indexes. If a punched card were made at the time of publication of each abstract or article, such cumulative indexes could be automatically compiled and printed in coordinate form within a few hours simply by sending the cards thru a computer as described above. Hence the cost of preparing such a cumulative index would probably be reduced at the same time as its value might be enhanced by becoming a complete coordinate index.

4. For survey of current foreign literature

It is often important just to know that an article on a particular subject exists. For example, it is claimed that the Russians published several articles about their Sputnik I before it was launched—of which the Western World was u naware. A periodical survey or list of current Russian scientific literature that included a coordinate index might have enabled the proper people to know of the existence of these articles even if the articles had not yet been translated. Such a survey and coordinate index is relatively inexpensive and can enable United States scientists to become aware of Russian articles on subjects of particular and specialized interest. Only the titles of the articles would be translated, and the coordinate index would of course be in terms of English words. A scientist, having found by means of the coordinate index a Russian article on a subject of particular interest, might then have it translated. Otherwise the article could have been entirely overlooked. In this way access can be had to hundreds of Russian journals that may contain important articles, but which, owing to their nature and various other circumstances, would in general *not* be translated at all.

5. In coordinating current research and development contracts

A bound book coordinate index of current research and development contracts, placed on the desks of research scientists could significantly aid the coordination of a large-scale research program, stimulate a more extensive and healthy information exchange between scientists, and prevent costly, time-consuming and unnecessary duplications. Up to now, large research and development programs, such as those of the Department of Defense, usually keep such indexes on punched cards at a centrally located installation; and they are interrogated upon request. However, such a central service has the serious disadvantage that a scientist must wait in line for his request to come up for processing, the request often involves preparing official correspondence by mail or memorandum, and the efficiency is low because the retrieval is done by someone other than the scientist himself. Such disadvantages can be eliminated by publishing periodical coordinate indexes in bound book form, and sending these to the contractors and research scientists.

6. In conjunction with open-ended card type coordinate files

Very frequently open-ended coordinate indexes on current periodical scientific or other literature are kept and compiled in the form of punched cards by organizations such as the Library of Congress. Each month many new articles or reports are received and added to the index. At the present great rate of research publication, such indexes can rapidly become too huge to handle. For example, an index of 100,000 reports may receive over 20,000 additional reports each year. Hence it is desirable to "close off" or take out of this current card index about 20,000 of the older reports each year. By means of the methods described above, whenever it is desirable to close off a section of such a current card index, a bibliography which includes the new coordinate indexing system can automatically be printed within a few hours simply by sending the cards directly through a computer. This bound book bibliography can then be mailed to libraries and laboratories throughout the country. In this way (1) the current open-ended punched card coordinate index can be maintained at a reasonable size, (2) past information can still be retrieved in coordinate index form from the bound book bibliographies, and (3) the compiling and printing of the bound book bibliographies can be completely and automatically accomplished from the cards within a few hours on a computer.

7. As a rapid method for machine retrieval

Searching literature by means of computing machines is an important aspect of information retrieval techniques. If the information is recorded in our tabular forms, great technical advantages can result for the retrieval process. These advantages stem from the fact that when one starts to retrieve from the tables, he effectively has a head start over other techniques. For example, suppose the article-word associations are recorded on magnetic tape, and the retrieval is carried out by searching the tape. Consider the following comparison of the effect on retrieval of three modes of recording the article-word associations on the tape (See Fig. 2.) First, suppose each article is grouped with its associated words as in (1) of Fig. 2; then in order to retrieve, *every group on the tape must be searched,* comparing the words of each group with the given words for retrieval, and those articles that are associated with at least all of the given words are recorded. Second, suppose each word is grouped with its associated articles as in (2) of the figure; then in order to retrieve, *the groups associated with each of the given words must be searched,* and the common articles recorded. Finally suppose the tables themselves were recorded on the tape, as in (3) of the figure. Here only a single table need be searched, namely the one

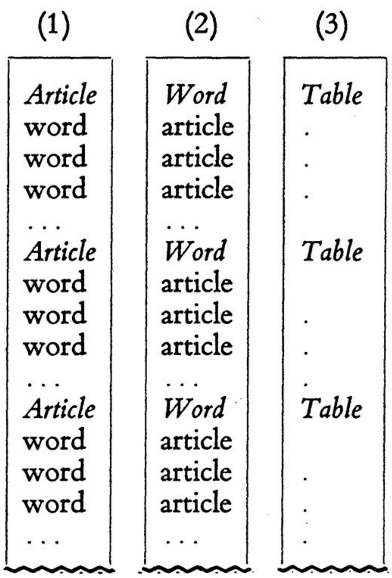


FIGURE 2. Three modes of recording information on magnetic tape.

corresponding to the first of the given words as described above. Depending upon the technical details of the computer being used, the fact that by means of the new method only one group or table need be searched can lead to significant advantages. Of course the other two illustrations presented do not exhaust the techniques for recording and searching magnetic tape, but they do serve to render a comparison to some extent.

5. ILLUSTRATIVE BIBLIOGRAPHY

PART I. THE ARTICLE LIST (BIBLIOGRAPHY PROPER)

PAGE 1

Article No.	
1.1	Abrahams, A.P. Autoradiographic determination of radioactivity in rocks. *Nucleonics* *15*:85–86 Mar 1957
1.2	Aravindakshan, C. A simple arrangement for obtaining optical transforms of crystal structures. *J. Sci. Instr. 34:* 250 Jn 1957
1.3	Gasstrom, R. V. A very fast pulse-height analyser with independent uptake, sorting and storage of information. *Nuclear Instruments* *1*:75–79 Mar 1957
1.4	NBS Circular 580. Bibliography on ignition and spark ignition systems. Nov 1 1956

PAGE 2

Article No.	
2.1	Nicholls, J. Alpha-scintillation monitor for hands and clothing. *Nucleonics 15*:80, 81, 83, 84 Mar 1957
2.2	Pope, M. I. An automatically recording vacuum balance. *J. Sci. Instr. 34*:229–232 Jn 1957
2.3	Powell, D. A. An apparatus giving thermogravimetric and differential thermal curves simultaneously from one sample. *J. Sci. Instr. 34*: 225–227 Jn 1957
2.4	Seidle, F. G. P., et al. Modification of the Brookhaven fast chopper. *Nuclear Instruments 1*:92–93 Mar 1957

PAGE 3

Article No.	
3.1	Senior, D. A. The Kerr cell, a high speed electro-optical shutter, Pt II. *Instr. Pract.* *11*:471–476 May 1957
3.2	Smith, B. O., and Grimshaw, A. G. A pneumatic level indicator. *Instr. Pract. 11*:469–470 May 1957
3.3	Stockendal, R., and Bergkvist, K. E. Evaporation device for beta-spectrometer samples. *Nuclear Instruments 1*:53–54 Jan 1957
3.4	Tove, Per-Arne. Electronic time analyzer applied to the measurement of the half-lives of metastable nuclear states. *Nuclear Instruments 1*:95–100 Mar 1957

PART II. THE WORD LIST (CODE BOOK AND THESAURUS)

PAGE 4		PAGE 5		PAGE 6	
	Word No.		*Word No.*		*Word No.*
adaptability	3.2	England	5.2	mass	1.2
analysis	4.1	evaluation	6.1	Netherlands	4.2
application	5.1	gas	2.2	nuclear	4.3
concept	6.2	heat	2.3	plan	6.2
counting	2.1	hysteresis	1.1	theory	4.4
design	7.1	implementation	5.3	thermal	2.3
differentiation	3.1	instrumentation	5.3	use	5.1
				versatility	3.2

PART III. THE INDEX TABLES

PAGE 7

Table 1.1

3.4	1.2 3.1 3.2 4.1 4.2 4.3 4.4 5.1

Table 1.2

2.2	2.2 3.1 4.1 5.1 5.2 6.1 7.1

Table 2.1

1.1	3.2 6.1
2.4	4.2 4.3 4.4 5.1 5.3 6.2

Table 2.2

1.4	2.3 5.3 7.1
2.2	3.1 4.1 5.1 5.2 6.1 7.1

Table 2.3

2.3	3.1 3.2 5.1 5.3 7.1
1.4	5.3 7.1

Table 3.1

3.4	3.2 4.1 4.2 4.3 4.4 5.1
2.3	3.2 5.1 5.2 5.3 7.1
2.2	4.1 5.1 5.2 6.1 7.1

PAGE 8

Table 3.2

3.4	4.1 4.2 4.3 4.4 5.1
2.3	4.1 5.1 5.2 6.1 7.1
1.1	6.1

Table 4.1

3.4	4.2 4.3 4.4 5.1
3.3	4.2 4.3 4.4 5.3 6.1 6.2
1.3	4.2 4.4 6.1
2.2	5.1 5.2 6.1 7.1

Table 4.2

2.4	4.3 4.4 5.1 5.3 6.2
3.4	4.3 4.4 5.1
3.3	4.3 4.4 5.3 6.1 6.2
1.3	4.4 6.1

Table 4.3

√ √2.4	4.4 5.1 5.3 6.2
√ √3.4	4.4 5.1
√3.3	4.4 5.3 6.1
2.1	6.2 7.1

Table 4.4

2.4	5.1 5.3 6.2
3.4	5.1
3.3	5.3 6.1
1.3	6.1

PAGE 9

Table 5.1

2.3	5.2 5.3 7.1
2.2	5.2 6.1 7.1
3.2	5.2 6.2 7.1
2.4	5.3 6.2
3.4	—

Table 5.2

3.1	5.3 6.1 6.2 7.1
2.3	5.3 7.1
1.2	6.1 6.2 7.1
2.2	6.1 7.1
3.2	6.2 7.1

Table 5.3

3.1	6.1 6.2 7.1
3.3	6.1 6.2
2.4	6.2
1.4	7.1
2.3	7.1

Table 6.1

1.2	6.2 7.1
3.1	6.2 7.1
3.3	6.2
2.2	7.1
1.1	—
1.3	—

Table 6.2

1.2	7.1
2.1	7.1
3.1	7.1
3.2	7.1
2.4	—
3.3	—

Table 7.1

1.2	—
1.4	—
2.1	—
2.2	—
2.3	—
3.1	—
3.2	—

APPENDIX I ILLUSTRATION ON AN ACTUAL-SIZE TABLE

The possible effectiveness of the method can best be seen by trying a retrieval on an actual-size table. Although it is difficult to estimate the size of an actual table, say for a bibliography of 10,000 articles and 6000 words, we have attempted to illustrate such a table in Fig. 3. Suppose it was desired to find all articles associated with the given word numbers (which have already been arranged in ascending order):

17.16, 24.03, 35.31, 53.26, 72.07, and 89.03

The first step is to turn to Table No. 17.16. Next the nonunderlined numbers of the rows are searched for the second given word number, 24.03. It will be noticed that all rows below the indicated arrow, being greater than the second given word number, need not be searched. Then the rows that contain 24.03 are searched for 35.31, and so forth. The appropriate checks as they finally look are illustrated. Only articles 136.03 and 32.02 are associated with *all six* of the given word numbers.

APPENDIX II FINDING ARTICLES ASSOCIATED WITH ALL EXCEPT ONE, OR ALL EXCEPT TWO, OR ALL EXCEPT THREE,…, OR ALL EXCEPT N OF THE GIVEN WORDS

It may often happen that, given a list of words, there exists *no* article in the bibliography associated with all of them. That is to say, the result of the above procedures on the index tables of a bibliography produce *no* articles that are associated with all of the given words. In such cases, one of two paths can be taken. The first is to relax the conditions, i.e., drop one or more of the words from the given word list. In many instances reconsideration of the purposes of the particular retrieval problem indicates that some of the given words are not as important as others and can therefore be dropped from the list. Of course, the smaller the list of given words, the greater is the chance that there exist articles in the bibliography associated with all of them.

An Estimated Example of a Typically Large Indexing Table

Table 17.16

for 89.03	for 72.07	for 53.26	for 35.31	for 24.03		
√	√	√	√	√	136.03	20.32 21.06 23.17 24.03 35.31 43.16 53.26 72.07 78.16 89.03
					27.05	20.41 21.01 23.01 25.32 37.25 52.22 72.07 77.15 78.16 81.08
				√̶	3.36	21.01 21.03 22.14 24.03 42.16 42.42 63.14 81.08 89.03
			√̶	√̶	106.21	21.03 21.10 24.03 35.31 42.16 42.42 78.16
					148.16	21.10 22.19 35.31 42.42 81.08
√	√	√	√	√	32.02	22.14 24.03 25.09 35.31 53.26 72.07 77.01 89.03
		√̶	√̶	√̶	76.15	22.15 23.17 24.03 35.31 42.16 53.26 75.10 76.19 81.08 89.03
					65.38	23.01 25.09 37.25 42.16 42.42 77.15 81.08 89.03 93.03
	√̶	√̶	√̶	√̶	137.43	24.03 25.09 35.31 42.16 53.26 72.07 76.19 77.01
				→	16.10	25.32 42.16 43.42 63.14 81.08 89.03 93.03
					118.08	37.25 42.16 63.14 77.15 89.03 93.03
					91.07	42.16 42.18 43.42 81.08 89.03 91.01 93.03 95.01 97.01
					24.73	56.02 63.14 77.15 81.08 89.03 93.03 95.01 97.01
					56.81	63.14 77.15 81.08 89.03 95.01 97.01
					75.22	77.15 81.08 93.03 95.01
					143.01	81.08 89.03 93.03
					151.83	93.03 95.01 97.01

FIGURE 3.

On the other hand, it may often occur that all of the given words seem equally important. In such a case it would be desirable to see if there are any articles associated with *all except one of the given words,* or if there are any articles associated with *all except two of the given words,* and so forth. In order to accomplish this, the second path is followed. The procedures for accomplishing this are described in this section. For example, suppose it is desired to find all articles associated with all of the following given word numbers (see Fig. 3):

17.16, 24.03, 25.09, 42.42, 72.07, and 81.08

It happens that there are no articles associated with *all six* of these given word numbers. To determine if there exist articles associated with all except one of these given word numbers, the straightforward method would be to successively omit one number at a time; this would require a total of seven "passes" through the tables: six passes omitting a number each time plus the first pass. It so happens that in the example there exist *no* articles in the bibliography associated with all except one of these given word numbers. Hence we would continue and try to find all articles associated with all except two of the six given word numbers; this would require fifteen more passes, or a total of twenty-two passes through the tables, which would certainly be a laborious process.

However, there is another scheme for producing the desired results. It turns out that by this new scheme only three passes are needed instead of the seven, and only six passes instead of the twenty-two. In fact, if there are n given word numbers, the total number of passes P needed to find all articles associated with at least $n-k$ of the given word numbers is:

$$P = \sum_{r=0}^{k} \frac{n!}{r!\,(n-r)!}$$

by the straightforward successive omission method but only

$$P = \frac{(k+1)(k+2)}{2}$$

by the new technique to be described.

Although the new scheme is rather complicated to describe in words, it is *exceedingly simple* to perform once the method becomes clear.

The new scheme can best be described in terms of the concept of a *pass,* which will now be described. A pass is associated with the indexing table of the smallest of a particular set of given word numbers. In the following discussion the smallest of the given word numbers is called the first given word number, the next to the smallest, the second given word number, etc. Hence, 17.16 is the first given word number, 24.03 is the second given word number, 25.09 the third given word number, etc., for the set of numbers in our example.

There is the first pass, second pass, third pass, etc., for a table. Hence the symbol (T 17.16/P2) is used to represent the *second pass* for Table 17.16, where 17.16 is understood to be the smallest of a previously given set of word numbers. The first pass for a table is a little different from subsequent passes, and we shall start by describing the first pass by means of an example.

The first step is to draw on tracing paper the same number of vertical lines as there are given word numbers, in our case six lines, and then overlay the paper on the indexing table of the first given word number, in our case 17.16, tucking the tracing paper into the binding of the book so that it will not easily slip.

These vertical lines are labeled on top by the successive given word numbers in order, starting with the second given word number. Henceforth these vertical lines will be called the *given word number lines* (see Fig. 4), and each line will be denoted by its label. Now search each row of the table for the second given word number. Draw a horizontal line from the second given word number line to the third given word number line corresponding to each of the *rows* that contain this second given word number. These horizontal lines will henceforth be called the *article row lines*. Next, in those rows that contained the second given word number, i.e., those rows which now correspond to an article row line, search for the third given word number; if it appears, extend the corresponding article row line from the third given word number line to the fourth given word number line. Next, in those rows whose article row lines have just been extended, search for the fourth given word number; extend the article row lines another segment, and continue in this method until no more article row lines can be extended. The work for our example, using the table in Fig. 3, is drawn in Fig. 4. Only rows corresponding to articles 136.03, 3.36, 106.21, 32.02, 76.15, and 137.43 include the second given word number 24.03 and hence article row lines are drawn corresponding only to these articles. Next, searching these rows, the rows of articles 32.02 and 137.43 are found to include the third given word number 25.09; the corresponding article row lines are extended. Searching these rows, the given word number 42.42 is not found; hence the procedure has been concluded. This is the process denoted by the first pass for a table.

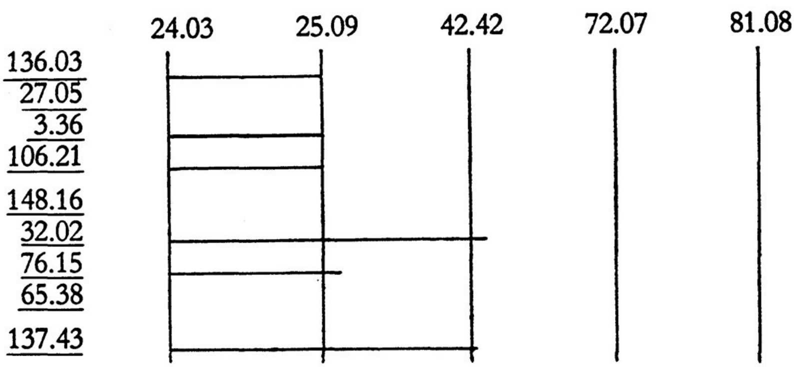


FIGURE 4. First pass.

Now consider a pass other than the first pass as, for example, the second pass at the table in Fig. 3. The first step is to relabel the given word number lines; this is done starting with the third given word number (see Fig. 5). Now, in those rows

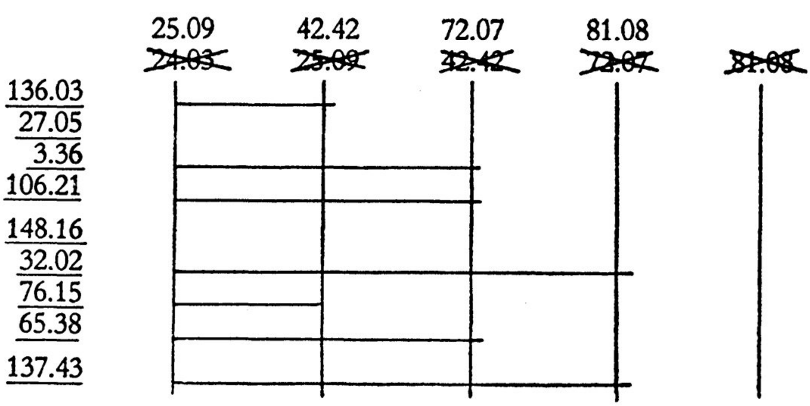


FIGURE 5. Second pass.

that do not have a corresponding article row line (i.e., those rows that did not contain the second word number), search for the third given word number and draw the corresponding first segment of the article row line where indicated from the newly labeled third given word number line to the newly labeled fourth given word number line. Next, in those rows whose article row line ends on this fourth given word number line, search for the fourth given word number; where it is found, extend the corresponding article row line another segment; and so forth, as before. The procedure for passes other than the first pass differs from that for the first pass only in that given word number lines must be relabeled, and initially only those rows that do not correspond to any previous article row line are searched. Note, of course, that in order to remember for which given word number any row is to be searched, one merely has to observe the label on the given word number line where the article row line ends (see Fig. 4 and 5). The work for the second pass appears in Fig. 5.

Figure 5 illustrates that first searching rows 27.05, 148.16, and 65.38 for the given word number 25.09, we find only row 65.38 includes the word number; and the article row line for *65.38* is extended accordingly. Next, searching rows 136.03, 3.36, 106.21, 76.15, and 65.38 for word number 42.42, we find that only rows 3.36, 106.21, and 65.38 include this word number. Then, searching rows 3.36, 106.21, 32.02, 65.38, and 137.43 for word number 72.07, we find that rows 32.02 and 137.43 include this word number. The article row lines are extended accordingly. Finally, searching rows 32.02 and 137.43 for word number 81.08, we find that 81.08 is not included in either row; and the process for the pass is concluded.

The process for a third or fourth pass is similar to the process for the second pass. The process for the third pass of our example is illustrated in Fig. 6.

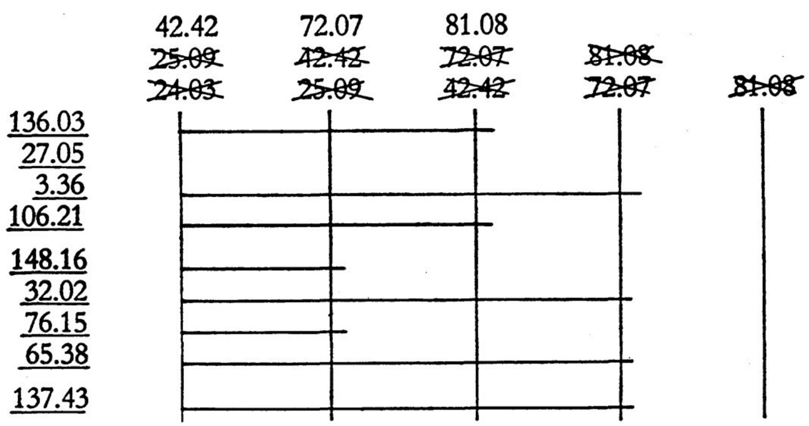


FIGURE 6. Third pass.

A pass is said to be brought to a *successful conclusion* if, when the procedure ends, at least one article row line has been drawn past the largest given word number line of that pass. The articles corresponding to these successful article row lines are called the *results of the pass.* For example, (T 17.16/P3) is brought to a successful conclusion, as illustrated in Fig. 6, where the results of the pass are 3.36, 32.02, 65.38, and 137.43.

Now we are ready to describe the new scheme. Let us consider as an example the table in Fig. 3 and the given word numbers listed in Fig. 3 (Table 17.16). The first step is to proceed with pass (T 17.16/P1). The lack of any results indicates that there is no article associated with *all* of the given word numbers, and so we proceed to find all articles associated with at least five of the six given words. To do this, we proceed with pass (T 17.16/P2) and pass (T 24.03/P1), where for the latter pass the set of given word numbers is understood to be the original set with 17.16 omitted. The results of these passes will be the desired answer. If again there are no results, then we would proceed to find all articles that are associated with at least four of the six given words. These will be found as the results of passes (T 17.16/P3), (T 24.03/P2), and (T 25.09/P1), where for the latter pass the set of given word numbers is understood to be the original set with both 17.16 and 24.03 omitted. The procedure now becomes clear. For instance, to find all articles that are associated with at least three of the six given words, we would list the results of the following passes: (T 17.16/P4), (T 24.03/P3), (T 25.09/P2), (T 42.42/P1). (See Fig. 7.)

All	(T 17.16/P1)			
All except one	(T 17.16/P2)	(T 24.03/P1)		
All except two	(T 17.16/P3)	(T 24.03/P2)	(T 25.09/P1)	
All except three	(T 17.16/P4)	(T 24.03/P3)	(T 25.09/P2)	(T 42.42/P1)
…	…	…	…	…

FIGURE 7.

The second and third passes are, of course, extensions of the first pass and even though they are here illustrated in three different figures, in practice the work will be done on the same "pass diagram."

APPENDIX III USING WORDS INSTEAD OF NUMBERS

The previous discussions were based on the use of numbers to represent the words. On the other hand, there is an alternative method, encompassing the same principles, that is based on the use of the *words themselves*. For this method, Part II (the word list) no longer associates numbers with the words. Instead, the word list just becomes a list of all possible words by which one can look up an article. Hence, step 1 simply reduces itself to making certain that the *given words* (by which it is desired to retrieve) are listed in the word list. Step 2 consists of putting the given words in *alphabetical order* and looking up the table associated with the *first* word. The tables are now labeled with the *words themselves,* there being one table for each word and the tables, accordingly, being placed in alphabetical order. Step 3 is similar to the original method except that words are found in the table where nonunderlined numbers were found before. The words are in alphabetical order within each row, and the first column of words is in alphabetical order.

As an example of the use of the words themselves, we have given on the following pages an illustrative bibliography based on the matrix of Fig. 1. Part I is, of course, the same as that given in Section 5, Part I. Part II has the synonyms and word numbers omitted. In Part III there are words in the tables instead of the word numbers. To find all articles associated with the application of nuclear theory, we put the three words in alphabetical order: *application, nuclear, theory*. Then the *application* table is turned to and *nuclear* looked up. It appears in two rows: 2.4 and 3.4. In these two rows, *theory* is looked up; it appears in both of these rows; so the desired articles are 2.4 and 3.4 by Seidle and by Tove.

The advantage of using words in the tables is twofold: first, looking up words is a familiar process in our society, and second, the use of words enables the user to make associations that he might otherwise not think of. If in each row of a table are listed *all* the words associated with the article of that row, the association aspect would be complete. In fact, in this case, the set of tables would closely resemble an ordinary index, but one which is exceedingly complete. Such a close resemblance to an ordinary index greatly simplifies the use of the tables as far as the user is concerned.

PART I. THE ARTICLE LIST

PAGE 1

Article No.

1.1 Abrahams, A. P. Autoradiographic determination of radioactivity in rocks. *Nucleonics* *15*:85–86 Mar 1957

1.2 Aravindakshan, C. A simple arrangement for obtaining optical transforms of crystal structures. *J. Sci. Instr. 34*:250 Jn 1957

1.3 Gasstrom, R. V. A very fast pulse-height analyser with independent uptake, sorting and storage of information. *Nuclear Instruments* *1*:75–79 Mar 1957

1.4 NBS Circular 580. Bibliography on ignition and spark ignition systems. Nov 1 1956

PAGE 2

Article No.

2.1 Nicholls, J. Alpha-scintillation monitor for hands and clothing. Nucleonics *15*:80, 81, 83, 84 Mar 1957

2.2 Pope, M. I. An automatically recording vacuum balance. *J. Sci. Instr. 34*:229–232 Jn 1957

2.3 Powell, D. A. An apparatus giving thermogravimetric and differential thermal curves simultaneously from one sample. *J. Sci. Instr. 34*:225–227 Jn 1957

2.4 Seidle, F. G. P., et al. Modification of the Brookhaven fast chopper. *Nuclear Instruments* *1*:92–93 Mar 1957

PAGE 3

Article No.

3.1 Senior, D. A. The Kerr cell, a high speed electro-optical shutter, Pt. II. *Instr. Pract. 11*: 471–476 May 1957

3.2 Smith, B. O., and Grimshaw, A. G. A pneumatic level indicator. *Instr. Pract. 11*:469–470 May 1957

3.3 Stockendal, R., and Bergkvist, K. E. Evaporation device for beta-spectrometer samples. *Nuclear Instruments* *1*:53–54 Jan 1957

3.4 Tove, Per-Arne. Electronic time analyzer applied to the measurement of the half-lives of metastable nuclear states. *Nuclear Instruments* *1*:95–100 Mar 1957

PART II. THE WORD LIST

PAGE 1	PAGE 2	PAGE 3
analysis	England	mass
application	evaluation	Netherlands
concept	gas	nuclear
counting	hysteresis	theory
design	instrumentation	thermal
differentiation		versatility

PART III. THE INDEX TABLES

Analysis

2.2	application	design	differentiation	England	evaluation	gas	mass
3.4	application	differentiation	gas	Netherlands	nuclear	theory	evaln.
3.3	concept	evaluation	instrumentation	Netherlands	nuclear	theory	
1.3	evaluation	Netherlands	theory				

Application

2.4	concept	counting	instrumentation	Netherlands	nuclear	theory
3.2	concept	design	England			
2.2	design	differentiation	England	evaluation	gas	mass
2.3	design	differentiation	England	instrumentation	thermal	versatility
3.4	differentiation	hysteresis	mass	nuclear	theory	versatility

Concept

2.4	counting	instrumentation	Netherlands	nuclear	theory
3.1	design	England	evaluation	instrumentation	
1.2	design	England	evaluation		
3.2	design	England			
2.1	design	nuclear			
3.3	evaluation	instrumentation	Netherlands	nuclear	theory

Counting

1.1	evaluation	versatility		
2.4	instrumentation	Netherlands	nuclear	theory

Design

2.2	differentiation	England	evaluation	gas	mass
2.3	differentiation	England	instrumentation	thermal	versatility
3.1	England	evaluation	instrumentation		
1.2	England	evaluation			
3.2	England				
1.4	gas	instrumentation	thermal		
2.1	nuclear				

Differentiation

2.2	England	evaluation	gas		
2.3	England	instrumentation	thermal	versatility	
3.4	hysteresis	Netherlands	nuclear	theory	versatility

England

2.2	evaluation	gas	mass
3.1	evaluation	instrumentation	
1.2	evaluation		
2.3	instrumentation	thermal	versatility
3.2	—		

Evaluation

2.2	gas	mass		
3.3	instrumentation	Netherlands	nuclear	theory
3.1	instrumentation			
1.3	Netherlands	theory		
1.1	versatility			
1.2	—			

Gas

1.4	instrumentation	thermal
2.2	mass	

Hysteresis

3.4	Netherlands	nuclear theory	versatility

Instrumentation

2.4	Netherlands	nuclear	theory
3.3	Netherlands	nuclear	theory
2.3	thermal	versatility	
1.4	thermal		
3.1	—		

Mass

2.2	—

Netherlands

3.4	nuclear	theory	versatility
2.4	nuclear	theory	
3.3	nuclear	theory	
1.3	theory		

Nuclear

3.4	theory	versatility
2.4	theory	
3.3	theory	
2.1	—	

Theory

3.4	versatility
1.3	—
2.4	—
3.3	—

Thermal

2.3	versatility
1.4	—

Versatility

1.1	—
2.3	—
3.4	—

APPENDIX IV POPULATION VS. SUBSTANTIVE MEANING OF THE TABLE ENTRIES

The entries in the tables can be chosen to convey information concerning the population of articles with respect to words, or else chosen to convey some substantive meaning associated with the words themselves. Thus in Section 5 of this paper, the entry numbers associated with the words described how many articles are associated with each word. On the other hand, in Appendix III, the words themselves were used in the tables, and as such the table entries conveyed the meaning of the words. In this appendix we discuss the relative advantages and disadvantages of the two approaches.

Using population based entries has the practical advantage of making tables with many rows have few columns, and tables with many columns have few rows. In this way adjacent tables are relatively similarly shaped which is of some advantage in making up page formats, and in addition tends to reduce the number of entries that need be searched. When many words are to be coordinated together, in general, it is most efficient to choose as the first word for the coordination the one which will *eliminate* the largest number of article possibilities. In the procedure presented above, this is precisely what is accomplished when the table corresponding to the smallest word number is looked up. Continuing, it is best to choose as the second word for the coordination that one which will *eliminate* the most articles from those possibilities that still remain. The probability of choosing as the second word the one that fills this requirement is maximized in general by choosing as the second word the one with the next smallest word number, and so forth. All this increased efficiency of search depends upon the fact that a population reflecting number is used for the table entries. Finally, there is a possible psychological advantage in using such numbers, for since the meaning of the number has no substantive value, the procedures to be accomplished might be carried out with more precision as a strict routine. On the other hand, when using entries with substantive meaning, one may be more tempted to deviate from the correct method by changing words in the middle, and this will give wrong results.

The disadvantage of using population based entries is that, if a bibliography is updated and republished including additional articles, the numbers for the words will change; i.e., the number for the same word will in general be different in the newer version than in the older version of the bibliography. On the other hand, if the entries in the tables have substantive meaning, the entry symbol corresponding to each word will be the same in any updated bibliography as it was in the original. This has the advantage that a person who has memorized a symbol corresponding to a substantive meaning of a word will not be thrown off when a bibliography is updated. In addition the substantive symbol may be more familiar to the user of the bibliography, as for example is the case in Appendix III where the word itself was used as its own symbol. Also when looking in a table, the substantive symbols found as entries might have *greater suggestive and associative* properties which is completely lost when population based symbols are used.

The kind of entry to be employed in any particular bibliography depends on the purpose and use of the bibliography. If the bibliography is small and if it is to be periodically updated and extensively and frequently used by the same individuals, then perhaps entries with substantive meanings should be used. On the other hand, if the bibliography is very large, and will not be revised, and if it will in general not be referred to so often by an individual that he will unconsciously memorize the entries for the different words, then the population based entry is probably most efficient.

ACKNOWLEDGMENTS

The author would like to acknowledge the encouragement and advice of Richard Dahl of the Office of the Navy Judge Advocate General, John Scherrod and Charles Gottshalk of the Library of Congress, and Mary Stevens, Joshua Stern, and Robert Elbourn of the National Bureau of Standards The author would also like to express his appreciation for support from the Information Systems Branch of the Office of Naval Research, and the Office of Scientific Information of the National Science Foundation.

The Comac: An Efficient Punched Card Collating System for the Storage and Retrieval of Information[1]

MORTIMER TAUBE

The concept of item codes and term codes has been fully developed in a paper by Mr. Alexander Kreithen which has also been submitted to the Conference (1); and from this paper we take the conclusion that there are only two basic patterns of grouping codes in a store: "Either term codes are collected under item codes or item codes are collected under term codes" (2). We also utilize the conclusion that, in a system of direct free field coding, a search consists of matching (or collating) a code (or codes) in the question successively against codes in the store.

Although these theoretical conclusions are generally applicable to any type of storage and retrieval device, in this paper their implication will be applied to the design of a specific system, namely, a new system of punched card collation which we have designated the *C*ontinuous *M*ultiple *A*ccess *C*ollator (Comac).

Since it is the purpose of this paper to demonstrate that the Comac is an efficient device for even the largest collections of information, e.g., patents, intelligence files, newspaper morgues, and picture files, we will base our calculations of search time and size of the store on the following figures:

Number of items in the store, 1,000,000.

Average number of term codes used to index an item, 20.

Number of terms in the vocabulary or different term codes used in the system, 10,000.

Within the general concept of matching we distinguish two methods employed by standard punched card devices. These two methods are usually shown as (1) searching and (2) collating.

MORTIMER TAUBE Documentation Incorporated, Washington, D.C.

[1] Prepared under Contract No. AF 49(638)–91, Air Force Office of Scientific Research, Directorate of Advanced Studies. October 10, 1957.

SEARCHING

Searching is performed with a *sorter* by making several successive sorts until all items coded by a certain term or terms have been selected, that is, sorted out from the rest of the deck. Changing the column selector in the sorter and selecting the cards from the proper pocket constitute setting up a question in the reading head of a piece of apparatus, and this question is matched successively against codes in the store. It is assumed that each item is represented by a card or set of cards on which is grouped the term codes characterizing that item.

With standard punched card equipment (unless superimposed punching or wiring is used) the term codes in the question must be matched against specified fields on the item cards. For example, a simple sorter "sorts" cards column by column as determined by setting the column selector in the sorter; and the "101" machine which can search many columns at once must be programmed to search in the proper columns for term codes in the question (the "101" can be wired to search for a single code in any of a number of fields).

The IBM Corporation has constructed an experimental searching device which can search for multiple codes any place on a card. This device, which uses direct free field coding, is known as the Luhn scanner; it represents a significant advance in punched card searching.

The operation of the Luhn scanner is illustrated[2] in Fig. 1.

Each item card in the store is coded with the terms characterizing the items and the unused columns of the card are "laced," that is, all holes are punched. The question card is also laced except for the columns required for the term codes in the question. The coding of a term in the question is the complement of the code of a term in the store (Fig. 2).

It is apparent that as a card in the store passes over the question card, the matching of complementary codes will cause a "blackout" which can signal a photocell to actuate a selection mechanism. The use of complementary coding and "blackout" cuts down the requirement for reading apparatus to one photocell per column per code area. If direct matching of codes were employed, each hole on the card would have to be read for a match or failure to match. (It will be seen that the use of complementary coding is only possible when a question card is prepared for a specific search and cannot be used in collation in which any card in the store may be used as a question card.)

[2] This illustration is only theoretically accurate. Since the scanner, as constructed, reads only one-half of the card, lacing is required on only one-half of the card area. Also, by virtue of always using codes of 5 out of 12 holes, lacing can be accomplished by just one extra punch which will always let light through.

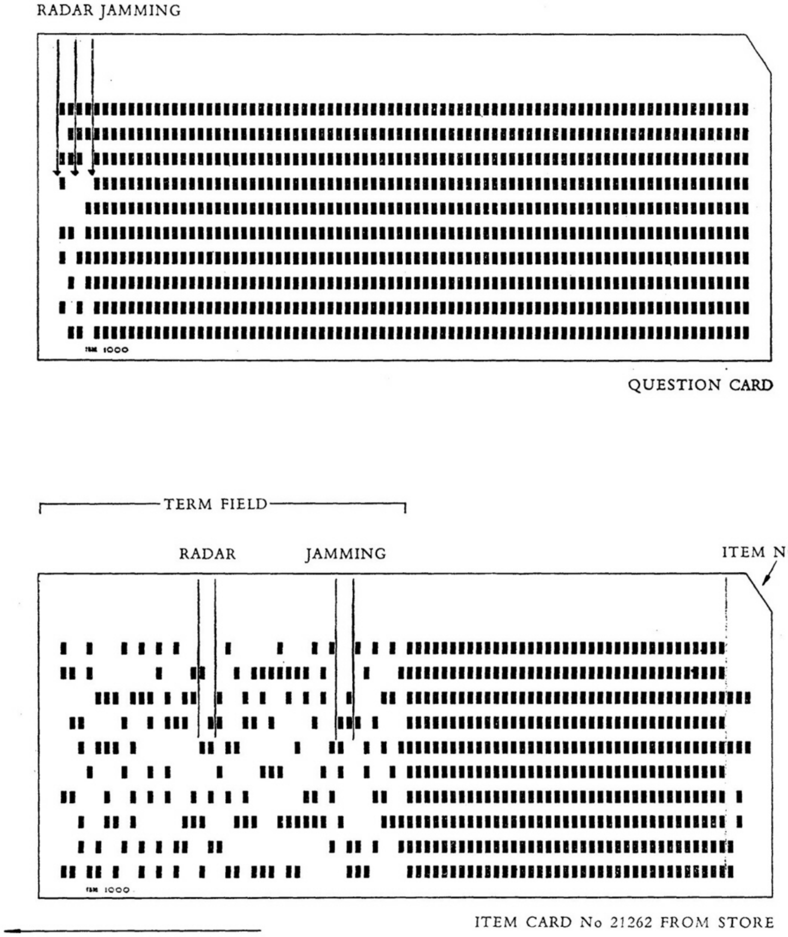


FIGURE 1

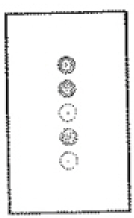


FIGURE 2. (*a*) Term code in question card.

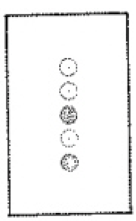


FIGURE 2 (*b*) Term code in store.

We will assume here that each card in the store to be searched by a Luhn scanner has room for twenty term codes and that no item is indexed by more than twenty terms. The Luhn scanner matches cards in the store against the question card at a rate of 1000 per minute. Since it is a necessary characteristic of searching systems (systems in which terms are collected under items) that the total file be scanned in any search, it would require 1000 minutes or approximately 16 1/2 hours to search a million items to answer one question. Hence a searching system even with so advanced a device as a Luhn scanner can only be used for relatively small collections or for collections which permit the division of items into mutually exclusive classes, each one of which is small enough to make searching the total class practical. The great advance of the Luhn scanner was its demonstration that free field coding could be used with punched cards and that one card could constitute the question which interrogated the store on other cards.

COLLATION

In spite of this development the inherent inefficiency of linear search has so far precluded the successful application of punched card searching to collections of any significant size which cannot be divided into mutually exclusive classes; but several relatively successful punched card installations have been organized for collating rather than searching. In setting up a system of punched cards for collating as contrasted with searching, grouping of items by terms is employed rather than grouping of terms by items. The following figure illustrates the two forms of grouping.

Searching

1	A	M	N	O
2	B	C	D	T
3	A	B	M	R
4	L	N	O	P
5	C	G	H	K
6	F	G	M	P
7	L	P	R	T
8	H	K	L	S
9	B	C	R	S
etc.				

Collating

A	1		3						
B		2	3						9
C		2			5				9
D		2							
F						6			
G					5	6		8	
H					5				
K					5			8	
L				4			7	8	
M	1		3			6			
N	1			4					
O	1			4					
P				4		6	7		
R			3				7		9
S								8	9
T		2					7		
etc.									

When collation is used as a matching technique, item codes collected under one term are matched against item codes collected under another term. In effect one group of item codes becomes the question which is matched against the other group considered as the store. It should be apparent that collation does not require the search of the total store but only of those item codes grouped under the terms of the question.

However, with standard collating equipment, a considerable price must be paid for this decrease in search time. A collection of 1,000,000 items indexed by an average of 20 terms would require a file of 20,000,000 cards. With 10,000 terms in the vocabulary the 20,000,000 cards would be arranged in 10,000 groups averaging 2000 cards in a group. Since a standard collator feeds 240 cards per minute from each feed, the collation of two terms (asking a two-termed question) would average between 10 and 20 minutes. This is an appreciable reduction from 16 1/2 hours, but there are some penalties which must be faced which reduce radically the efficiency of standard collators as information searching devices.

In the first place the size of the store must be increased enormously to permit prefiling items (cards) under every term by which they are indexed, in this instance, from 1,000,000 to 20,000,000. Secondly, collators work only on arrays maintained in fixed numerical or alphabetical order. Hence, item cards must be filed (posted) to each term array and maintained in that array in a fixed order. Thirdly, cards matched by the collator and selected as answers must be refiled in proper order. If the selected cards are to be retained as an answer or are to be matched or are to be matched against existing groups, they may have to be duplicated so that the array from which they are selected initially can be restored to completeness for other searches.

These difficulties, which arise from the use of standard collators as information searching devices, are not attributable to the grouping of items by terms, but to the use of a collator designed primarily for interfiling of cards rather than the matching of codes. It will be seen that most of the difficulties disappear when a device like the Comac is substituted for a standard collator.

The standard collator carries out its interfiling function by noting the match, the failure to match, and the order of numbers on cards which it compares. On the basis of what the match discloses, the collator advances one deck or the other, or both (in the case in which a match occurs). The multiplication of cards from one to twenty million is not attributable to the need for additional coding space but to the fact that, since the collator reacts to a match by selecting cards or to the recognition of order by interfiling cards in proper order, it can operate on only one item code per card. But if we separate the collator's ability to match codes from the requirement for physical selection and inter filing of cards, it becomes possible to put more than one item code on a card and to signal a match by punching the matched code on another card.

THE COMAC

The essential function of the Comac which determines its design is simply the ability to match codes on one punched card against codes on another punched card and to punch the codes for the logical product or sum on a third card. Consider a set of item codes on card A and another set on card B. (See Fig. 3.)

The card AB can be collated with card C, etc. The final answer if it involves the product [(A∪B)∪C)] can be printed rather than punched.

It is immediately apparent that one of the features of the Comac is the fact that it does not require any refiling of selected cards into an A deck or a B deck. The degree of file reduction, however, may not be immediately apparent.

An IBM card contains 80 columns. Since, in collation, we group item codes under term codes, let us assume that 2 columns are required for the term code of each card and 3 columns for the card number. With one column blank, the remaining 74 columns can be divided into 37 two-column groups with each group containing 24 bits. The 24 bits can be divided into 6 fields of 4 giving a possibility of coding any item number from 1 to 999,999. Hence 37 item numbers ranging from 1 to 999,999 can be punched in each card.

In terms of our previous figures, namely, 1,000,000 items, 20 term codes per item, 10,000 terms in the vocabulary, the 20,000,000 item codes could be punched on 540,540 cards, providing a file reduction of 37 to 1 as compared with standard collating systems.[3] In more picturesque terms the card file is reduced from building size to room size. The 540,540 cards would be organized into 10,000 groups averaging 54 cards to a group. And the collating of one group against the other would involve comparing the item codes of 54 cards with the item codes on 54 cards. The groups will ordinarily not be equal and many groups will contain many more than 54 cards; but unlike the Minicard System, we can add cards to any group without dedicating space for it in the group. Hence it is appropriate to discuss the groups of a Comac system in terms of averages.

It is interesting to note that the 3,000,000 patents in the Patent Office could be handled on approximately 1,600,000 cards and given the same 10,000-word vocabulary, with an average of 162 cards in a group.

[3] This form of binary coding in columns is called "Chinese Binary" by IBM data processing people. Although it gives maximum compression of codes, it does require decimal binary converters and modification of existing punches. With regular Hollerith coding, the 37 to 1 file reduction possible with Chinese binary, drops to 14 to 1 for collections requiring 6-digit codes (up to 999,999 items) and to 18 to 1 for collections requiring 5-digit codes (up to 99,999 items).

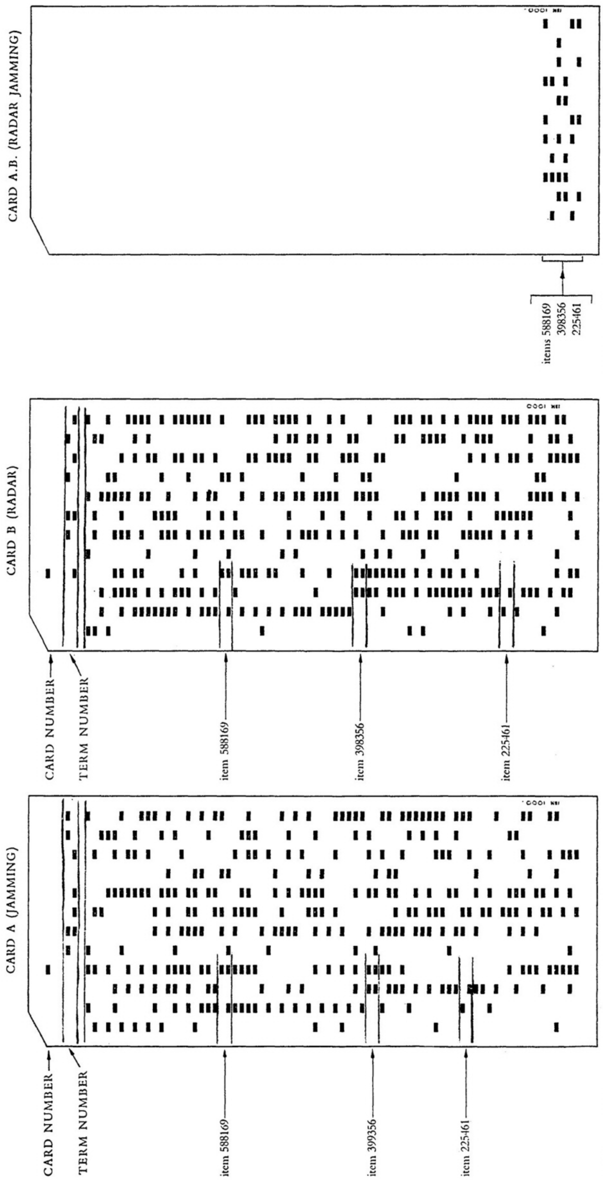

CARD A.B. (RADAR JAMMING)
items 588169
398356
225461
CARD B (RADAR)
CARD NUMBER
TERM NUMBER
item 588169
item 398356
item 225461
CARD A (JAMMING)
CARD NUMBER
TERM NUMBER
item 588169
item 399356
item 225461

FIGURE 3

Once the basic design requirement of the Comac is established, the Comac can operate in either of two ways depending on the amount of comparators or registers that are provided. If only one code on each card can be read at a time, both groups being collated would have to be advanced intermittently as is the case with existing collators, the difference being that the cards would be advanced two columns at a time instead of a card at a time. The Comac which compares only one code at a time from each group can only operate if the item codes are punched in ascending order, that is, if "code 1< code 2< code 3, etc." This is also the case with existing collators. This constraint can be avoided if enough circuitry is provided to store all the codes on a card while another card is passed over it. In such a case all codes would be matched against all codes regardless of the order of the codes on the cards. Whether or not it would be worthwhile to provide this additional circuitry would depend on the nature of the collection being indexed. If the items in the collection could not be assigned serial numbers and indexed in order, the more advanced type of Comac would be required; but if serial order could be maintained in indexing and in entering item codes on cards, the Comac which advanced and compared one code at a time from each group would be adequate.

We have not attempted in this paper to describe the Comac apparatus. However, from our studies of existing punched card equipment, binary to decimal converters, comparators, etc., it appears that once the basic concept of the Comac is accepted, the construction of a device for single code comparison represents only a very modest development effort. Actually the character of the physical equipment necessary is practically deducible from the new concept of collation as a matching and print-out process of item codes rather than a card selection and interfiling process. The development of the more advanced Comac capable of comparing multiple codes with multiple codes might require a greater investment; but before such a development is undertaken, the value of removing the constraint of entering item codes in order should be thoroughly explored.

ADVANTAGES OF THE COMAC SYSTEM

We summarize at this point the characteristics of the Comac which make it a practical and efficient information storage and retrieval system. Some of these points have been mentioned above; others will be presented here for the first time.

1. DECREASE IN TIME OF SEARCH

For a collection of 1,000,000 items, the grouping of item codes under term codes, as compared with term codes under item codes reduces the time of search for a two-termed, question from 16 1/2 hours (Luhn scanner) to approximately 5 minutes. We are assuming that cards can be advanced two columns at a time and compared in the Comac at about twice the rate they can be advanced in existing card reproducers that feed cards the long way, i.e., one card every 2 to 3 seconds. Since one card in the Comac contains 37 codes, 54 cards containing 2000 codes can be read in 100 to 150 seconds. Doubling this figure to allow for the intermittent advance of two groups, we get 200 to 300 seconds or 3 to 5 minutes.

2. MULTIPLE ACCESS TO THE STORE

Searching involves scanning the whole file; but collating involves comparing prefiled groups. This means that many searchers can select groups for comparing at the same time without interfering with one another. Further, it is assumed that the Comac will be so reasonable in price that any large and busy installation would have several so that searchers desiring to collate term groups would not have to queue up at one machine. Assuming 5 minutes for the average search, five Comacs would provide an answer every one minute during a working day.

3. ELIMINATION OF REFILING

The card reproducing and print-out features of the Comac would eliminate the necessity which exists with present collators of refiling cards. That is, there would be no “return-to-normal” problem which now exists with most punched card searching and selection devices. For example, the Patent Office R & D Report No. 6 describes this problem as a major difficulty in the utilization of punched cards for searching (3).

4. CARD REPRODUCTION AND PRINT OUT

The Comac will be able to reproduce a punched card containing matched codes; but in addition it will contain a binary to decimal converter which will print out the final answer of a succession of collations, e.g., $\{[(A \cap B) \cap C] \cap D\}$.

5. EASE OF MAINTENANCE AND POSTING

The cumulation of item codes for punching on term cards is a simple procedure which has been worked out by Documentation Incorporated and other organizations (NSA) in connection with posting on Uniterm Cards (4).

6. FREEDOM FROM CONSTRAINTS ON INDEXING

The reproducing features of the Comac make it fairly simple to combine terms, set up hierarchical relations between terms and the items grouped on such terms, etc. For example, a decision to establish a grouping of item codes under the general term antibiotics, that is, to collect codes previously listed under aureomycin, streptomycin, penicillin, etc., involves only changing instructions in the reproducer. Whereas an ordinary search involves the reproduction of the logical product of the group of item codes, the collection of item codes under a general term is equivalent to reproducing the logical sum of the codes.

Hence, although it is possible to look upon the Comac as a mechanized Uniterm Index, the mechanization extends not only to the comparison of codes but to the ability to update the actual indexing and to provide any degree of order or hierarchical relationship required by the search problems confronted by the system.

REFERENCES

1. KREITHEN, A., “Mathematical Foundations for a Storage and Retrieval Theory,” AFOSR IN 57–400. ASTIA AD 132475, JUNE 1957.
2. Reference 1, p. 14.
3. DON ANDREWS. U.S. Patent Office, Research and Development Report No. 6, pp. 11–12.
4. SANFORD and THEREAULT. Problems in the Application of Uniterm Coordinate Indexing. *College and Research Libraries, 17,* No. 1, 19–23 (January 1956).

SUMMARY OF DISCUSSION

INTRODUCTION

The task of the discussion panel, as outlined by Dr. Gilbert King, was to try to present a coherent picture of the status of developments in information storage and retrieval systems, developments which are in a state of flux. The subject is not yet a science, and there was no attempt to classify the submitted papers in a formal arrangement. Rather, the discussion was divided into two pieces, called the “what we have to do” section and the “how to do it” section.

To a scientist trying to make use of information, such questions arise as, are there really any good retrieval systems, and do scientists in the laboratories really use these systems? Dr. King said that one cannot point to a system which has appeal and stimulation and which has been spontaneously adopted.

One reason is that in present systems the process of putting material into the store or library can be considered a translation process, from the material language of the text into the index language of the system. The user is forced to ask his questions in this primitive language. We must introduce some kind of syntax, so that scientists can communicate with libraries.

The devisers of systems seem to think that scientists do their work by asking specific questions, but this is not so. What the scientist wants from the library is suggestions, parallelisms, and stimulation. To get this, he must be able to ask questions freely. But as already noted, he must learn this peculiar language invented for him by librarians. So one big problem, not discussed in the papers in this area, is an operations analysis of the type of questions which are asked of a library. How can the users learn to speak in the language employed to organize the information in the store?

In Dr. King's opinion, we shall not make much more progress until we have a rigorous mathematical model of storage and retrieval systems, whether mechanical or manual. In the case of the simpler indexing and classification schemes the model is very simple. Each document is assigned a binary number. Each digit corresponds to some description or index head. The problem of classification is merely a question of how to arrange an arbitrary set of binary numbers. Retrieval is merely the question of searching for binary numbers with “ones” in specified locations. More importantly, we realize immediately that the language of the library and the language that the interrogator is required to use are both the simple language of binary numbers.

Some papers discuss a model of retrieval in which the binary numbers are arranged as a partially ordered set, but even these models reveal the lack of texture in the organization. The relationship of documents can only be expressed by the limited properties of binary numbers. Their environment is not at all subtle and therefore can hardly be expected to provide sophisticated search procedures. It is extremely important that more complex retrieval systems be invented, and formulated in a similar kind of mathematical model.

TYPES OF QUESTIONS

It is ten years since there was a similar conference and in most disciplines of science there are no more information services available for the scientist now than there were ten years ago. One of the main reasons for this, according to Dr. Dyson, may be that we are thinking on too broad a basis. He said that it would seem advisable to devote ourselves to the solution of a certain number of smaller practical problems in the various disciplines, rather than trying to cover the whole realm of scientific thought.

One of the first things to settle is what kinds of questions do we want to answer in new systems, which present systems do not enable us to answer. These questions would be different in the different disciplines. In organic chemistry, for example, we cannot now find all the organic compounds of iodine in regular indexes, without reading through the whole index. This kind of search would be useful if it could be carried out fairly easily.

One reason why this type of search has not been made possible is that we have not yet made a sufficient study of language. Nobody has yet ascertained whether a basic language in organic chemistry can be created, to simplify the question of putting vast amounts of knowledge into mechanical records.

Another type of search we would like to make in organic chemistry is the correlative search, where one searches for substances with certain common properties. The language of new information systems must be designed so that these searches can be made.

Mr. Perry emphasized that we must not fail to consider the importance of the questions to be asked and the form these questions might take. Answering questions is the objective in any system, and economic benefits result from the ability to answer questions. It is the margin of benefit over costs that will be decisive in the selection and operation of new systems. But we must beware of false assumptions that impede or misguide efforts to emphasize the practical side in the development of documentation systems.

One assumption is that the use of the literature is not much influenced by the available facilities for getting at it. But even the questions posed in the library are strongly influenced by the available means for getting the questions answered. Many persons ask questions in terms of what they think a conventional subject index can answer, rather than what a machine searching system can answer. To design our new systems on the basis of the kind of questions that now are directed to the library is to overlook what we can do. Of course this does not suggest that we should abolish abstracts and indexes because there will always be questions for which such tools will be extremely useful.

The forms of questions asked may be very diverse, Mr. Perry said. Often, people ask questions with no idea of what they really need. In mechanized systems, a preliminary search can be run to orient oneself in a field. At the present time and probably for the foreseeable future, we shall be operating in terms of first identifying what is of probable interest. This involves relating the questions to the means established for getting answers: looking under the proper subject headings in an index, using the right class numbers in a classification scheme, devising instructions for programming in a mechanized system. By using the wrong headings or class numbers, or devising a wrong program, one can run a search and report that nothing on the subject exists in the system. This kind of error will not be changed by mechanized systems.

FUNCTIONS OF INFORMATION SYSTEMS

Suppose, now, we were suddenly confronted with the ideal retrieval machine so that the mechanics of retrieval is easy—would we know what to do with it? In Dr. Swanson's opinion, the benefit of thinking in terms of such a machine would be the ability to separate the conceptual and intellectual problems from the strictly hardware problems. Then, at some point in the future we will be able to tell the machine designers what is actually wanted in the way of a machine.

We do not know whether the operations we now know how to perform with machines would lead to a better system than the present methods of getting information out of libraries. The operations we now know how to perform can be enumerated: we can search for words; we can give the machine a dictionary and search for synonyms and near synonyms; we can search for logical combinations of words and require that they occur in certain proximity to one another. Even granting that we could do all this, what would be our procedure for evaluating such a system? No soundly designed experiments have been carried out to measure the efficiency of retrieval.

One of the critical questions in evaluating a retrieval system, Dr. Swanson stated, is what still remains in the library of interest that was not recovered by any of the known methods of retrieval. We may have a false sense of security that we have retrieved everything just because we get a large response to our question.

We now have the tools for carrying out experimental observations on what we would do once we are presented with the hypothetical ideal machine. Once we study the retrieval process we may decide to keep the separation into indexing and then retrieval operations. It might well be more economical to perform once certain operations which are identical to all questions, and reserve for the retrieval processes those operations which vary from question to question.

In considering the practical aspects of the subject, Dr. Sanford commented that it is interesting to note the competitive nature of the problems discussed in the papers in this area. The scientist requires a fast, exact information service. He also requires a service which will provide bibliographic search. Unfortunately, the best organization of material to serve these two basic needs is different in each case. In the storage of information, choices also have to be made; terms are collected under item codes or item codes are collected under terms. We have the added rivalries of subject matter. Various kinds of scientific subjects require different approaches in analysis and organization. We now inquire as to what to do, and this is presumably the question for which this conference was called. The members of this panel are sure of this, that an integrated effort will grow and mature only in a climate of active cooperation among us.

Dr. Sanford went on to suggest that in seeking an answer as to what to do, we might consider system compatibility as the first objective. We already have satisfactory answers to a large segment of our problems, but only when each problem is considered alone. When we try to go from one individually good system to another, we find we lack a bridge. Language itself might be used as the missing bridge. Let us explore this use of more direct language. The need for compatibility is one of our most immediate and fundamental problems. This compatibility must be not only between disciplines but also among the tools for handling information. We must also consider the need for economy, not only economies in cost but also in effort.

Many electronic developments are ready to serve us as soon as we know how to use them. Considering machines as additional tools of our trade, their first contribution may be to provide the missing links between our systems. We might be able to continue to use our own separate classification schemes without the penalties of insularity. Machines might also provide the economies of handling which we need. The evolution of regional, national, and international centers and clearinghouses equipped with electronic devices already existing, could provide bridges between our present specialized disciplines and methods. These centers could bring order out of our present kaleidoscope of compatible problems, if we are willing and able to think and work together.

Mr. Cleverdon said that in establishing cooperation between librarians and those who come into documentation from other fields, the librarians have the right and duty to tell what their problems are and to direct the efforts of these specialists toward the solutions of the problems. The librarians must make sure first that the problems are not of their own making. For example, many work in the vast field of the applied sciences. It appears probable that the useful life of the information in these collections is reasonably short. The time for retaining references in catalogs should be equally short. The librarians should make certain they are not creating their own difficulties by allowing subject catalogues to become large and complicated. Only when all irrelevancies are cleared out can the true problem be posed to those who are trying to help.

We already know enough of the capabilities of machines in retrieval to know that within the limits of the information put into them and if correctly programmed, they can do everything a human can do in a clerical sense, but we are justified in asking what will be the cost. And we can doubt that machines will for many years to come solve all the problems in information storage and retrieval.

For example, Mr. Cleverdon continued, one comment on the papers in this area is that many appear to be fascinated by the ultimate goal and are unwilling to appreciate the efforts required at the input end. Our really hard work is in indexing. This is a chore which should be lightened as far as possible. The indexer has to decide on the subject content of the document. Then the chosen subject must be translated into the language of the indexing system. More work should be done from the input side, for the work of indexing is the most important subject of information retrieval. The more straightforward this task can be made, the more efficient and satisfactory will be the retrieval.

Dr. O'Connor focused his comments on questions that have been raised about certain difficulties in using mechanical systems. The Crane-Bernier paper, for example, raised these questions about manipulative search systems in general. Certainly subject indexes are easily accessible compared with access to a machine documentation center. Searching in a printed index or through a card catalog has certain virtues in keeping one in immediate contact with the subject matter. But many of the problems suggested for manipulative systems can be solved. Thus it is possible to avoid the danger of using too many search terms in a question, and thereby not finding relevant papers, by a device called sub-searching. By this one searches for all documents containing all the terms one specifies; those containing all but one of the terms, those with all but two of the terms, etc., to those documents containing just one of the terms. Sub-searches may be done automatically on the IBM 101 machine or on computers. Otherwise by making a number of separate searches these sub-searches are accomplished. They are also a way of getting related information while looking for relevant documents. The paper by Opler and the work of the Cambridge Language Research Unit and the U.S. Patent Office all touch on this question of relevant browsing in a mechanical system.

Dr. O'Connor said that other questions raised as to the difficulties in using mechanized systems, such as objections as to programming time or delay in getting information, can similarly be answered when one is more experienced in the use of machines, what they can do, and how they can be used to good advantage. As for costs, we do not have enough cost information now to know whether a mechanized search costs more than a non-mechanized search. Even if it should cost more, the information found may be worth it.

CLASSIFICATION AND INDEXING

Research on classification going on in England was presented in some papers. Mr. Farradane said that the workers there define classification as the ordered representation of an idea or a complex of concepts in such a way that a unique symbolization of it is given in an ordered sequence and it is retrievable because it has a position which can be defined. That is, classification is not only a hierarchical division; it can be indexing as long as the index is sufficiently ordered to give a sequence with a unique and recognizable order.

Facet analysis divides a subject into a number of different groups of terms of the same kind. Then ideas are expressed by reassembling these terms in various combinations with considerable freedom. This method has been proved a useful tool in various practical cases in England. Within a special field of knowledge a special classification can be made along faceted lines which gives a considerable degree of flexibility and usefulness in handling the field. In putting terms in a preferred order, one assumes some reason for that order which is really an implied relation between the terms. This problem, of relational connection between ideas, is a field of study which requires a great deal more attention. All this work requires more basic research, basic experimental investigation of what we are dealing with and what we are trying to do.

When asked what is the ordering relationship practiced by the British workers, Mr. Farradane replied that inclusion is one of the possible relations; others are hard to define but the preferred sequence gives the implied relations between terms.

Dr. O'Connor commented on the differences between some British and American approaches to retrieval systems. In America we tend to have an unorganized list of descriptive terms, or a list organized only to permit the user to search it to find the terms he wants to use in his question. In the British systems, the list of index terms is so organized as to impose an order on the documents to which those terms are assigned. The intent is that the order imposed on the document collection will be useful for searching.

In answer to a question as to how to handle relations with the peek-a-boo retrieval system described in his paper, Dr. Gardin reported on three ways of expressing them. A simple solution is to use inflected terms, specifying whether a term is used as subject or object in the action. That gives a first degree of abstraction useful when you do not have to take into consideration the exact type of action taking place. If you want to indicate what happens between the subject and the object you have to add a third term, usually a verb.

To indicate verbal relations, the type of action actually taking place, you can use two very general positions, he said, such as positive action and negative action. There is a third type of relational term which concerns topographical relations, indicating terms such as “to the left,” “above,” and “interweaved.” These three classes of terms are treated just as if they were entities as a being or noun in a lingual system.

USE OF LARGE-SCALE MACHINES

In our efforts to organize information, we have again not yet taken full advantage of present capabilities of automatic data-processing machines in the analysis of material previous to even starting a system. Mr. Luhn pointed out that we can employ machines to great advantage to do this analysis, to give us material on which to build a sound system. We can use the machine to encode this material. A searching procedure may be tested by setting up experiments where a computer simulates conditions under which a system will operate and tests it to find optimum conditions of operation. This is a saver in time and money for designers of retrieval systems.

We want to assign to machines the run-of-the-mill chores. As we get more experience we can give the machine more and more of what we have to do. There is, however, a residual which may never be turned over to the machine. This is the place where the librarian plays the role of interpreter. The interrogator approaches the librarian, gives him the questions and ideas, and the librarian then communicates with the machine.

Another point brought out by Mr. Luhn was that in our attempts to mechanize indexing and retrieval we are imitating systems designed originally for human capabilities. Machines have entirely new capabilities. Unless we utilize them to the fullest, we will never solve this problem. It will take many years before we learn what machines are and how they would prefer to act, but this is the general direction in which we will have to go.

According to Dr. Moore, in our discussions we can find three levels of what machines can do: present day capabilities; capabilities after machines have learned to read printed text; and capabilities after machines have learned how to think. Misunderstandings develop unless we are quite specific as to which level we are talking about.

Present day capabilities require us to do a lot of translating ahead of time into special code. Systems in use at the present time are based on a large amount of clerical transformation and critical analysis of the information to be put into the system. There are some problems definitely amenable to such machines. A good example is the organic structural formula problem, of identifying and searching for compounds with certain fragments, Dr. Moore said.

Again the concordance type of analysis, making an exhaustive list of every occurrence of every word, is something that is feasible and easy for a machine to do.

Dr. O'Connor noted that there are ways in which machines can help in the production of a traditional search tool, a printed subject index for example. Or a computer suitably programmed might take an input of index entries and determine what would be the most efficient and the most easily searched arrangement of the index entries.

Machines that can read are possible, in Dr. Moore's opinion, but their application to scientific material will pose some problems. Thus if you want to read *Chemical Abstracts* you need to have over 900 symbols at your disposal. A device to read these would be much harder to build than one that can read 50 or 100 symbols.

The machines that “think” are logically possible but whether or not they are economically feasible is another matter. If we can build a machine that can think and still keep it interested in doing drudgery and the bibliographic jobs we now assign to machines, then we will be on an entirely different level of discussion than we are on now.

In the area between the two levels, of machines to which you give explicit codes and those which can read, there is an interesting character recognition problem, that of recognizing bibliographic references—to decide whether or not two references are to the same document. Dr. Moore observed that if the ingenious idea of a citation index is to be carried out economically, it would depend on the machine taking bibliographies of the technical papers, sorting them internally, and comparing to see which are the same references. Then the machine could type out an inverse listing, giving you all the articles which cite a given document. It would seem that this is an easier recognition problem than reading ordinary letters.

On the general subject of mechanized indexing and abstracting, even though it might not be feasible on present computers in terms of cost, then it might very well be feasible on the next generation of computers, according to Dr. O'Connor. One of the questions that occur on reading of Harris' and Luhn's work is how to handle synonyms. A purely formal frequency count might be thrown off as to which are main subject words. And leaving some adjuncts together with the centers might throw off the frequency determination of the main subject somewhat. Mr. Luhn's thesaurus approach would take care of the synonym problem.

Two further complications arise from a mechanical index. Some articles might deserve as an indexing term a word not contained in the article. By an authority list, the product of the mechanized indexing procedure might have such additional words added to it. Again, an article might use a particular word but the vocabulary of the system might prefer another one. This also can be handled by a mechanized authority list.

Another complication is that mechanized indexing finds in a paper what was important to the author. What happens if there is something in the paper not important to the author but of importance to the indexer? One possibility is to have a list of words and phrases expressing the interests of a particular collection, which the machine looks for in the papers. If this word or phrase occurs even once, it should be picked up as an indexing term.

RESEARCH ON NATURAL LANGUAGES

"Why was indexing invented in the first place?" asked Mr. Luhn. It was so that we would not have to scan everything to recognize those things of interest to us. But machines can scan at fantastic speeds today and theoretically do not need indexing systems. (This is the basic assumption of automatic abstracting.) If we can take the inquiry in the open language of the inquirer and by means of the proverbial black box, translate this inquiry into the many variations that might occur in the stored text, then we could ask the machine to lead us to similar passages in the collection.

Dr. Swanson suggested that basically, we should give more attention to the storage of and operations on raw natural language texts. The separation into indexing and then retrieval becomes artificial if the capability of our machine is great enough. This does not mean we should abandon study on the organization and classification of documents. But in proportion far less work is now being done on how to handle raw text or natural language. What do we now know about the potential of operating upon natural language?

We should conduct more research on natural language text without regard to the division of input and output. A number of papers in this area were addressed to this problem, in particular the papers of Harris and of Masterman *et al.*

A number of people allege, Dr. Swanson continued, that there exists an analogy between mechanical translation and information retrieval. We can say that they both take as their starting point operations on natural language.

In the Harris paper linguistic transformations are discussed from the point of view of using them for the purpose of more compact storage or in the indexing process. They might first be used on natural language recorded in a computer in order to discover the relationship and usefulness of such transformations in the retrieval process.

The paper by Masterman *et al.* purports to discuss the question of analogy between mechanical translation and information retrieval. Dr. Swanson asked if one can translate from a foreign language into a pidgin English, then improve the translation in the target language by means of the thesaurus, why it is not easy to translate from English to English using the thesaurus? It would seem to be a worthy objective to establish before attempting translating from a foreign language into English. He also asked if, in translating from one language to another via an interlingual thesaurus, this thesaurus would have to be at variance with respect to language. That is, the thesaurus headings would have to be clearly identifiable in the source language and in the target language, even though the list of synonyms that appears under each would be different in the two languages.

Dr. Swanson went on to say that both mechanical translation and information retrieval take as their starting point operations on natural language and they are both somewhat concerned with the synonym problem. They might have one more attribute in common, namely that there may be certain approaches (e.g., the thesaurus approach) which fail for one as for the other.

In replying to Dr. Swanson's questions, Miss Masterman said that he is asking that the thesaurus should become a philosopher and should engage in a continual redefinition. And this process of redefinition within a field is just about the most sophisticated, profound, and subtle operation of which the human mind is capable. To require this of what must be at this stage a very primitive thesaurus seems to be philosophically unreasonable. About the other criticism of the interlingualness of the thesaurus, Miss Masterman felt that mechanical translation and the study of natural language cannot contribute to library retrieval unless we can find an interlingual way of doing it.

On the other hand, Dr. Bar-Hillel, of The Hebrew University, Jerusalem, said the hope that analysis of ordinary language would yield anything in the foreseeable future for information retrieval or language translation is totally unjustified.

Partial replacement, at least, of two-stage information retrieval by one-stage retrieval should be investigated, according to Dr. Bar-Hillel. We now go to the literature for reference to documents in which the information we need might conceivably be found. There is advantage in going straight to the information and finding out which sentences in the storage systems are relevant to our problem. But this would require first a transformation of the information from ordinary language into some kind of normalized system. The difficulties of doing this are reduced by the fact that in using scientific language our vocabulary is reduced, and the syntax of scientific writing is essentially less complex than that of ordinary expression.

The U.S. Patent Office is tackling its problem in the non-chemical field by a system of “ruly English.” Mr. Andrews of the Research and Development Division there pointed out that there are many complex relationships which cannot be handled by the syntax useful in the chemical field. The Patent Office is trying to build up a basic understanding of the language we use and speak every day.

By and large, natural languages are with us to stay, as Dr. Oettinger of Harvard said. A long-term view of the information problem requires acquiescence with the retention of natural language in these devices. It is also clear that artificial languages are here to stay inside machines. There are compelling reasons of technology and economy that dictate that inside a machine is a language best for the machine even though not best for humans. Here we are faced with a coupling problem.

The use of kernels as a means for generating abstracts or for picking out significant passages in a text, is all well and good once the kernel is formed. What would be of tremendous importance so far as the direct use of natural language in this area is concerned, is an effective algorithm for generating kernels.

This is all one area where some of the research on automatic translation may join with the problems of retrieval in that this problem (of finding kernels) is also one of automatic translation.

SOME TECHNICAL REQUIREMENTS

Papers in this area have not really covered the whole problem of the design of an information retrieval system, according to Mr. Buckland of Itek Corporation. The user has been left out, or has been paid only lip service all along. The user is an active rather than a passive element. Current machine design should try to make use of his tremendous inductive powers and use the interrogator in a feed-back loop. It might amount to giving the man more eyes and ears so that he might have data presented to him for him to make the decisions as to whether the data are relevant or not.

The author of an Area 5 paper, Professor Meredith, also observed that one of the stages in the process of communication, that of the user, has not been considered very much in this conference. Even those who have stressed the importance of the user of information have tended to assume that they are only concerned with those users of information who are specialists. There are also problems of the comprehension of scientific literature and its transmission across disciplinary boundaries. There are obvious conceptual difficulties and tools are needed which can help to analyze concepts into general terms.

Another author, Dr. Vickery, said that one purpose of this Conference is the improvement of the flow of scientific information at all levels. Although the prospects held out by the use of large computers are fascinating and exciting, we should not neglect the study of smaller systems. It is important to study not only how to design larger systems and more efficiently, but also to examine to what extent we can improve the efficiency of all the many small existing systems. The catalog based on faceted classification is admittedly only suitable for a relatively small system. The faceted method of analysis is an attempt to formalize the language, to control the terms and relations between them in constructing an indexing system of this relatively small scale.

Improvement of retrieval at all levels and all sizes of systems needs to be studied, he added, not merely to concentrate on what may indeed prove to be the hope of the future, that is, the direct use of natural language text.

Dr. Moore remarked that there are hopes that when one can read the full text of material it will not be necessary to use a concordance. The method of Mr. Luhn, of choosing the most frequent words and selecting sentences containing clusters of them as a way of writing abstracts automatically, is a first attempt in this direction.

Innovations like machine abstracting and machine indexing provide us with a facility which can give a significant statistical sample for analysis. Mr. Estrin, author of an Area 5 paper, suggested that Mr. Luhn should see that there is widespread distribution to abstractors and linguists of the products of the automatic abstracting process, so that they might vigorously attack the deficiencies. If he should get such feedback, he would get clues as to how to make whatever improvements can be made in his particular system.

SCIENTIFIC WRITING

Dr. Ranganathan concluded the panel discussion by considering the problems of scientific writing. One important factor we might lose sight of is the quantity of literature being produced, that has to be stored and retrieved. The scientist would prefer to have very few papers and very short ones. But why should so many different papers be produced on the same subject? We must examine whether there is any valid reason for repetition. We must examine the kind of audience as well as the kind of writing necessary to reach the audience. We can recognize six levels of audience and therefore six levels of writing in which any idea or any new thought should be embodied.

There is the master mind that creates ideas and expresses them in ordinary language. His audience is comprised of the tops among the intellectuals. Their job is to exploit what this genius has discovered. They have no time to write for the benefit of other people. They write for their peers. That is the second level of writing.

A third level communicates facts and data. This kind of writing is needed largely by engineers, technologists, and so on. We also want the knowledge to reach the people. It should be socialized. For that purpose we want somebody who can write to catch the attention of the people.

There is a need for another kind of writing. We want research to run in series, not in parallel, and we do not want unnecessary repetition. Therefore, we want report writing. We also want the ideas to reach down to those less mature, both in the physiological and psychological sense.

We want all of these types of writing, Dr. Ranganathan continued. But should we put all of them into storage for retrieval? This Conference is concerned only with the first three classes: seminal writing, knowledge writing, and data writing. But in storing them we have to select, and selection implies rejection. Unfortunately, this Conference has not thought about these techniques. But we should take them into consideration if we are going to make a store of reasonable size.

Dr. Ranganathan then asked, where classification comes in relation to machinery? In data writing, the data, facts, formulae, properties, are in associative-form. It is quite easy for these data to be fed into machines and to be retrieved as and when we want. Classification will have to do the facet analysis accurately and make it available to be fed into the machine.

In knowledge writings there are two difficulties. We cannot index them merely by words because the messages that they contain are often found not exactly in this word or that, but between the lines, so to speak. If we miss them we miss the document altogether. Machinery might not be able to spot these ideas. The other difficulty has to do with the words we use. When new thoughts emerge, we have no words prepared for them. So we must ask some existing words to take care of the new ideas. This raises the question of the possible efficiency of machines in indexing or abstracting these knowledge writings.

With regard to data documents, we can probably surrender some of the work to machines. Whether we can do the same with knowledge documents is questionable. We at least have to put into the machine some work already done by the human mind. That is the work of classification, Dr. Ranganathan concluded.

Professor de Grolier remarked that there were a number of papers in Area 5 which fall in the field of relations between linguistics and semantics and information retrieval. There are two different views on the point. One thesis seems to be that the present state of the language of science is more or less permanent. There is another school which tends to prove that there is something to be changed in the language of science from the point of view of better information retrieval. Professor de Grolier suggested that it would be useful after this Conference to have a conference on linguistics and semantics in relation to information retrieval.

MADELINE M.BERRY, *Rapporteur*
GILBERT W. KING, *Discussion Panel Chairman*

AREA 6

ORGANIZATION OF INFORMATION FOR STORAGE AND RETROSPECTIVE SEARCH

Possibility for a general theory

AREA ORGANIZATION

Authors of Papers

Members of Discussion Panel

Chairman: JOHN W.TUKEY, Department of Mathematics, Princeton University, Princeton, N.J.

YESHOSHUA BAR-HILLEL, The Hebrew University, Jerusalem, Israel

JULIAN BIGELOW, The Institute for Advanced Study, Princeton, N.J.

R.M.FANO, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Mass.

I.J.GOOD, Government Communications Headquarters, Cheltenham, England

R.A.KIRSCH, Applications Engineering Section, Data Processing Systems Division, National Bureau of Standards, Washington, D.C.

DONALD A.MACKAY, Department of Physics, Kings College, University of London, London, England

BENOIT MANDELBROT, Faculty of Sciences, University of Lille, Lille, France

MARVIN L.MINSKY, Lincoln Laboratory and Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Mass.

W.J.TURANSKI, Remington Rand Univac Division, Sperry Rand Corporation, Philadelphia, Pa.

PROPOSED SCOPE OF AREA 6

THE IMPORTANCE OF formulating a general theory to describe systems for the storage of information and subsequent search for retrieval justifies an investigation of the assumption that storage and search systems have enough in common to enable the development of a general theory. Without such a theory our efforts are directed to the empirical description of different devices which happen to get developed.[1]

But when we attempt to develop a general theory, our research is directed toward the discovery of common characteristics, common methods of functioning, common purposes, and common types of logical organization. When these common notions have been isolated and defined, it should be possible to account for the difference between the various systems and devices in quantitative rather than qualitative terms.

It follows that the first work to be done is the identification of a group of concepts which have a common relevance to all storage and retrieval systems. As examples of such concepts we can suggest: *A,* relatedness or connectivity of documents in a collection along with a suitable metric to define the degree of relatedness; *B,* storage space requirements; *C,* number of dimensions of access to a file; *D,* access time; *E,* coding efficienty. Hence it would be appropriate, in Area 6, to have papers that isolate and define such common parameters of storage and retrieval systems. It would be highly desirable if reasonable models of retrieval systems could be set up and analyzed, and if similarities between these models and models of different kinds of systems that have already been studied could be exploited. In this way it might be possible to make use of techniques that already exist in such studies as linguistics and information theory.

It is to be hoped that the search for quantitative relations among the basic concepts and the achievement of a theory of storage and retrieval systems will be able to predict new properties of such systems. If this can be done, any attempt at experimental verification of these properties would be an appropriate topic for papers, as would be any experimental investigation of the reasonableness of the fundamental postulates of the theory.

[1] The proposed scope of the Conference Area, as shown here, was prepared during the Spring and Summer of 1956 and provided to all potential contributors as a guide to the aims of the Conference.

Thus we have three sub-areas for papers in this general area:

1. Papers concerned with identifying, describing, and relating the basic concepts in the field of storage and retrieval theory.
2. Papers concerned with formal techniques of analyzing such systems.
3. Papers concerned with the systematic mathematical development and experimental verification of the theory. Perhaps it may be too ambitious to expect papers in the third area until work has gone forward in the other two.

The Structure of Information Retrieval Systems

B.C.VICKERY

Four basic operations in the effective use of graphic records (documents), to store information and make it available, have been listed by Hyslop: A, recording information in documents; B, storing recorded information—documentary items; C, identifying items containing information relevant to a given problem, situation, or subject; D, providing the identified items from storage. Information storage and retrieval in the wide sense covers all these operations. In the narrow sense used in this paper, *information retrieval* means only C, identifying documentary items by subject.

An *information retrieval system* is therefore defined here as any device which aids access to documents specified by subject, and the operations associated with it. The documents can be books, journals, reports, atlases, or other records of thought, or any parts of such records—articles, chapters, sections, tables, diagrams, or even particular words. The retrieval devices can range from a bare list of contents to a large digital computer and its accessories. The operations can range from simple visual scanning to the most detailed programming.

A retrieval system stores *units of information*. Each unit is linked in the system to specifications of one or more documents or parts of documents—I will call them *items*. The user specifies particular units of information—*specific subjects*— and the system is designed to provide him with a knowledge of all relevant items recorded in the system.

A retrieval system can be studied at three levels:

I. The way in which units of information, and relevant relations between them, are defined in the system. This is the semantic level of *subject analysis*.
II. The general structural features of the system considered as a network of units of information linked to each other and to documentary items. This may be called *structural analysis*.
III. The physical mechanisms (hardware) in which the structure is embodied.

A general theory of information retrieval would cover *(a)* principles of sub

B.C.VICKERY Imperial Chemical Industries Limited, Welwyn, England.

ject analysis, *(b)* a general study of structure, and perhaps *(c)* general experimental conclusions about the use of certain hardware. For two reasons, the present paper is restricted to *(b)*. Firstly, problems *(a)* and *(c)* seem properly to belong to Areas 5 and 4, respectively, of the Conference. Secondly, the technique of subject analysis, on the one hand, and the experimental study of physical mechanisms, on the other, do not appear yet to have advanced sufficiently for firmly based general conclusions to be drawn.

This paper, then, deals only with the general structural features of retrieval systems.

THE CHARACTERISTICS OF SPECIFIC SUBJECTS

Specific subjects are products of thought, and are typically in language or some other expressive symbolism. The units of meaning in such symbolism I will call *terms*. Without adducing formal proof, the following characteristics of specific subjects can be stated as facts of observation:

1. Each specific subject is symbolised by a single term or a combination of terms (*simple* and *compound* subjects respectively).
2. The same combination of terms can have more than one meaning, e.g., “the destruction of bacteria by dyestuffs” and “the destruction of dyestuffs by bacteria,” in other words, the terms within a specific subject are interrelated in certain ways.
3. Terms are also related to each other in at least two other ways: *(a)* by one term having a meaning which *includes* that of another term (generic relation) and *(b)* by both terms having meanings which are included in that of a third term (*coordinate* relation).
4. Specific subjects are consequently related to each other in at least three ways: *(a)* by each containing the same term or combination of terms, *(b)* by one specific subject containing a term which is generic to a term in another, and *(c)* by one containing a term which is coordinate with a term in another.

If T_1, T_2, T_3, and T_4 are terms, subject S_1 is said to *include* subject S_2 (i) when $S_1=T_1 \cdot T_2$ and $S_2=T_1 \cdot T_2 \cdot T_3$, or (ii) when $S_1=T_1 \cdot T_2$ and $S_2=T_4 \cdot T_2$, where T_1 is generic to T_4.

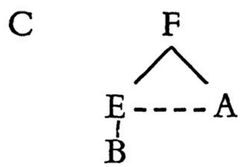
THE INFORMATION LATTICE

An information retrieval system includes a store of units of information, specific subjects. The assembly of specific subjects so stored may incorporate all the relations mentioned above. Between terms in each specific subject and between specific subjects, semantic links exist. An intricate network of links is thus formed, a *lattice* of units of information.

The links in the lattice are of three general kinds: *(a)* between terms in each specific subject there are *interlocking* relations, which may be represented by dots in the lattice, thus $T_1 \cdot T_2$, or by the multiplication sign, thus $T_1 \times T_2$; *(b)* between specific subjects there are firstly *inclusion* relations, which may be represented by solid lines in the lattice, thus S_1 —S_2, or by the inequality sign, thus $S_1 > S_2$; and *(c)* between specific subjects there are also *coordinate* relations, which may be represented by broken lines in the lattice, thus S_1- - -S_2, or by the equality sign, thus $S_1 = S_2$.

We can now state what is meant by items "relevant" to a particular sought subject. The limits of *relevance* can be varied at the discretion of the designer of the retrieval system. The system can be made to retrieve items recorded for the subject sought, and, in addition, items recorded for subjects (1) which include the subject sought, (2) which are included by the subject sought, and (3) which are coordinate with the subject sought.

Let us consider a small sample retrieval system. It uses six terms, words, which we can represent as the capitals A to F. Of these, C, D, and F are independent, E and A are included in F, and B is included in E. The lattice for these relations is as follows—by convention, a solid line leads *downwards* from the inclusive to the included term.



From these elements, ten specific subjects are derived. I will assume here that each subject coincides with a single documentary item, although this need not necessarily be the case. Each item is represented by a roman numeral.

Item	*Subject*	*Item*	*Subject*
I	B · C · D	VI	A · C · D
II	C · D · F	VII	A · D · E
III	B · C	VIII	D · E
IV	C · E · D	IX	A · B
V	A · C · E	X	A · D

The information lattice may now be extended to include these subjects, as in Figure 1. The complexity of even this small lattice (in which not all the coordinate relations have been drawn) is very evident.

Such a lattice is the structural basis for the store of an information retrieval system. The use of such a three-dimensional network as a retrieval system is, however, quite impracticable. The system must be simplified by breaking many of the links and reducing the lattice to a two-dimensional or linear structure. This simplification may be made outright, without taking any compensating action, in which case relations within the system, represented as links in the lattice, are lost. The result is that the complexity of retrieval is reduced, the scope of possible "relevance," as defined earlier, is narrowed. Alternatively, transformation to a system of fewer dimensions may be accompanied by compensating *marking* which preserves the lattice relations intact.

THE INFORMATION MATRIX

The first transformation may be to convert the lattice into a two-dimensional *matrix,* the coordinates of which are items and terms (Table 1).

TABLE 1.

	Items										
Terms	I	II	III	IV		V	VI	VII	VIII	IX	X
A (<F,=E)						x	x	x		x	x
B (<E)	x		x							x	
C	x	x	x			x	x				
					ⓧ						
D	x	x		x			x	x	x		x
E (<F,=A,>B)						x		x	x		
					ⓧ						
F (>E, A)		x									

The terms A to F are listed alphabetically, and the links between them are thus broken. In compensation, against each term its inclusion and coordinate relations must be shown. Thus A is included in F and is coordinate with E, so the solid line A—F is replaced by A<F, and the dashed line A- - -E is replaced by A=E. All these lattice relations between terms are thus preserved.

The remaining inclusion and coordinate relations are represented by the *postings* (here, crosses) in the body of the matrix. Consider subject A · C · D (item VI). Subjects which include it are derived by suppressing terms one by one: A · C, A · D, C · D, A, C, D. Search of the matrix reveals that only one of these subjects (A · D) refers to a documentary item (X) which may be retrieved.

Coordinate subjects are derived by replacing each term in turn, and search retrieves items I(B · C · D), II(C · D · F), IV(C · E · D), V(A · C · E), and VII(A· D · E). Since A=E, then A · C · D=E · C · D, so that item IV is again on a subject coordinate with that of item VI. Since A<F, then A · C · D<F · C · D, and the subject of item II includes that of item VI.

Lastly, there is the problem of interlocking relations within subjects. In the sample matrix it is assumed that these are only significant in one item, IV. Here, the subject in the lattice, C · E · D, is considered to differ in meaning from a subject in which C and E are not interlocked, i.e., C · D · E. In all other three-termed subjects, it is assumed that no significant interlocking occurs. The link between C and E in item IV must be recreated in the matrix by modifying the postings (here, doubling the crosses).

Having thus preserved all the lattice relations, the information matrix can in principle be used as a two-dimensional retrieval system. However, if a large number of terms and/or items were covered, the matrix would be intolerably large, a very high proportion of it would be wasted empty space, and the system would be very inefficient. The matrix must be further transformed.

The next transformation may be to partition the matrix either horizontally or vertically into units which can be inscribed on individual *entries* or *tallies*. Horizontal partition produces entries each bearing a single term and its associated items. Vertical partition produces entries each referring to a single item, bearing the terms associated with it. Partition into units greater than a single row or column is possible, and indeed necessary in certain cases.

The two-dimensional matrix is thus broken down into a series of units of the following types:

1. Term entry: A(<F,=E): V, VI, VII, IX, X
2. Item entry: I: B(<E), C, D
3. Interlocked term entry: C×E(<F,=A,>B): IV
4. Combined item entry: II+III: B(<E), C, D, F(>E, A)

The fourth type will not be further discussed.

The entries are inscribed with *marks* symbolising the terms and items. Information retrieval is the operation of selecting entries which bear marks for certain terms, and reading off their item specifications. The selection process thus operates on the term symbols.

The entries can be stored in the system in an order which is not related to the term symbols they bear, an arrangement which will be described as *entries free*. The selection of entries bearing a given term symbol can then only take place if the entire index is inspected. Systems so designed are described by Fairthorne as *scanning* systems. Alternatively, the entries can be stored in an order dictated by the ordinal values of the term symbols inscribed on them. Selection can then proceed by successive steps of “distillation,” and all entries need not be inspected. Such arrangements will be described as *entries fixed,* the result being a *segregation* system (Fairthorne).

TERM ENTRY SYSTEMS

A series of term entries derived from the previously tabulated matrix is as follows; the word *heading* referring to the combination of term symbols on each entry.

Index No. 1 (N_1)

Heading	*Items*
A(<F, = E)	V, VI, VII, IX, X
B(<E)	I, III, IX
C	I, II, III, IV, V, VI
D	I, II, IV, VI, VII, VIII, X
E(<F, = A,>B)	IV, V, VII, VIII
F(>E, A)	II

The only lattice relation which has been broken by partition is the interlocking between C and E in item IV. The relation can be restored by including in Index No. 1 an interlocked term entry, C×E(<F,=A,>B): IV. Search operations can then proceed as described for the matrix.

The construction of Index No. 1 requires only the following schedule:

Schedule No. 1 (L_1)

A(<F, = E)
B(<E)
C
D
E(<F, = A,>B)
F(>E, A)

The system thus consists of Index No. 1 (N_1) and Schedule No. 1 (L_1). The operations of construction are (1) look up words in L_1 to ascertain cross references, (2) inscribe headings on entries, (3) add interlocked term entries. The search operations are (1) consult sought words in N_1 (e.g., A, C, and D), (2) compare items to find those which are on all three entries (none are), (3) move to related entries by suppressing terms (A · C, A · D, etc.) and as indicated by cross references (e.g., F · C · D includes A · C · D) and repeat comparison in each case.

In this system, the marks symbolising terms are words, and inclusion and coordinate relations are expressed as verbal cross references. These features have two disadvantages: *(a)* Words may be long, far longer than is necessary to distinguish between them; if the schedule covers N terms, and r separate marks are available, then the symbol length need not exceed $n=\log N/\log r$ marks. *(b)* The use of words leads to alphabetical scattering of related terms, thus the coordinate terms A and E are separated, as is F from its subordinate A. This leads to complication of the search operation in systems with entries fixed. To overcome these disadvantages, other means of representing terms and relations may be adopted, leading to other schedules, indexes, and systems.

The simplest transformation is to replace the words A to F by *random code symbols* U to Z, which are no less "alphabetically scattered," but which are reduced to the minimum length needed to distinguish them. The resulting Index

No. 2 is identical with N_1, except that A to F become U to Z. A different schedule must be used in constructing the system.

Schedule No. 2 (L_2)

A	U(<Z, = Y)
B	V(<Y)
C	W
D	X
E	Y(<Z, = U,>V)
F	Z(>Y, U)

The second transformation is to minimise the distance between related terms. It is accomplished by a transformation into linear form of the lattice of relations between terms, and the allocation of *class numbers,* as shown in the following schedules.

Schedule No. 3 (L_3)

Class No.	*Word*
1	C
2	F
3	E
4	B
5	A
6	D

Schedule No. 4 (L_4)

Class No.	*Word*
1	C
2	F
21	E
211	B
22	A
3	D

In L_3, the independence of the three terms C, F, D is shown by their standing vertically in line, at the left of the schedule. Inclusion relations (F—E, E—B, F—A) are shown by *indentation,* coordinate relation (E- - -A) by vertically lining up. The class numbers simply fix the order of the terms, they are purely *ordinal*. In L_4, the independence of C, F, and D is shown by their being represented by unitary class numbers (1, 2, 3). Any term (e.g., B, 211) is included in any other term whose symbol can be derived by suppressing marks from the right-hand end of B, thus 21,E and 2,F. Coordinate relation is shown by two terms (E,21 and A,22) having common initial marks (2). As well as having an ordinal function, such class numbers are *hierarchical*.

From such schedules, indexes such as the following can be constructed.

Index No. 3 (N_3)

Heading		*Items*
1	or 1	I,II,III,IV,V,VI
1x3	or 1x21	IV
2	or 2	II
3	or 21	IV,V,VII,VIII
4	or 211	I,III,IX
5	or 22	V,VI,VII,IX,X
6	or 3	I,II,IV,VI,VII,VIII,X

In order to construct and use such indexes, a further schedule is needed, to lead from the words of the subjects indexed and sought to the symbols used in the system, thus:

Schedule No. 5 (L_5)

Word	*Ordinal No.*	*Hierarchical No.*
A	5	22
B	4	211
C	1	1
D	6	3
E	3	21
F	2	2

The use of ordinal class symbols is possible only in systems with entries fixed, and is further restricted to systems which can indicate relations by structural features other than the symbolism, as by indentation in N_3. Typical examples of such *static* indexes are the conventional card catalogue or the printed bibliography, in which purely ordinal symbolism is thus possible.

On the other hand, if the only device for indicating relations is the symbolism itself, the class numbers must be hierarchical. Such hierarchical class symbols have a function also in scanning systems with entries free. Although they do not then have the effect of juxtaposing related entries, yet they may be a simpler device for following up inclusion and coordinate relations than the original cross-references.

ITEM ENTRY SYSTEMS

An index of item entries derived directly from the previously tabulated matrix is as follows:

Index No. 4 (N_4)

Heading	*Item*
A(<F, = E)· B(<E)	IX
A(<F, = E)· C· D	VI
A(<F, = E)· C· E(<F, = A,>B)	V
A(<F, = E)· D	X
A(<F, = E)· D· E(<F, = A,>B)	VII
B(<E)· C	III
B(<E)· C· D	I
C·D·E(<F, = A,>B)	IV
C·D·F(>E,A)	II
D·E(<F, = A,>B)	VIII

Such an index can be constructed from Schedule No. 1. Once again, the only lattice relation lost is the interlocking between C and E in item IV. The relation can be restored if the terms in each heading are cited, not in a random or (as here) alphabetical fashion, but in an *interlocking order*. Thus the heading for item IV should become $C \cdot E(<F,=A,>B) \cdot D$. The heading as given in the index represents a subject significantly different in meaning. Such a solution is only possible in a system in which the terms occupy fixed relative positions in the entry, an arrangement which may be called *terms fixed.*

It is possible, however, to take a further step in transforming the original lattice, and to break the interlocking relations between terms (represented by dots in $C \cdot E \cdot D$). The positioning of terms on an entry then becomes random (e.g., they may be superimposed), an arrangement which may be called *terms free*. Two ways of restoring interlocking relations seem possible. *(a)* Thet firs is to include in the heading, not only the individual terms C, D, and E, but also a new compound CE. *(b)* The second is to modify the interlocked terms by the addition of symbols which indicate their linkage, thus C★, D, E★. For a system with terms free, therefore, N_4 must be modified in one of these two ways.

Further transformations of N_4 may be made exactly analogous to those of N_1, replacing the words A to F with random code symbols, to give N_5 (identical with N_4 except that A to F become U to Z), or with ordinal or hierarchical class numbers, as follows:

Index No. 6 (N_6)

Ordinal No.	*Hierarchical No.*	*Item*
1.2.6	1.2.3	II
1.3.5	1.21.22	V
1.3.6	1★.21★.3	IV
1.4	1.211	III
1.4.6	1.211.3	I
1.5.6	1.22.3	VI
3.5.6	21.22.3	VII
3.6	21.3	VIII
4.5	211.22	IX
5.6	22.3	X

Once again, the inclusion and coordinate relations must be shown by a non-symbolic feature (indentation) with ordinal class numbers, which are therefore only applicable in a system with entries fixed. Moreover, ordinal class numbers can only carry out the function of ordering related entries if the individual symbols in each heading (e.g., 1, 2, and 6 in item II) are taken in a fixed order, i.e., if the system has terms fixed as well as entries.

Hierarchical class symbols in a system with entries and terms fixed serve the same function, without the need for indentation. If the entries are free, the symbols cannot do this, but serve to indicate inclusion and coordinate relations. If the terms are also free, an interlocking device such as the asterisk must be added.

Access to item entry systems with entries and terms free is direct. Either the sought word, say D, its corresponding random code symbol derived from $S_2(X)$, or the hierarchical symbol from $S_4(3)$, is matched against the index entries. Since terms are randomly arranged in the headings, the whole of each heading must be scanned.

Access to item entry systems with entries and terms fixed is not so direct. In each heading, the order of terms is fixed, and direct access is only through the *primary* term. Thus direct entry to N_4 with D would yield only item VIII, and would not retrieve items VI, X, VII, I, IV, and II, in which D also figures. Access to fixed item entries therefore requires one of the following two alternative devices: *(a)* The entry may be partitioned into mutually exclusive *fields,* and any given term is always inscribed in a particular field. The entries can still be arranged so as to juxtapose related headings, but a search for all entries bearing a given term necessitates scanning the appropriate field in every heading. *(b)* No fields may be used, but an auxiliary *chain index* may be constructed in which each of the secondary terms in each heading is given a leading place in turn. Thus for index N_4 the chain index would be as follows:

Chain Index No. 1

B·A see A·B	D·C·B see B·C·D
C·A see A·C	E·C·A see A·C·E
C·B see B·C	E·D see D·E
D·A see A·D	E·D·A see A·D·E
D·C see C·D	E·D·C see C·D·E
D·C·A see A·C·D	F·D·C see C·D·F

For N_5, symbols U to Z would replace A to F after "see" in each line. For N_6, the following would be constructed.

Chain Index No. 2

		Ordinal No.	Hierarchical No.
A·B	see	4.5	211.22
A·C	see	1.5	1.22
A·E	see	3.5	21.22
A·E·C	see	1.3.5	1.21.22
B·C	see	1.4	1.211
D·A	see	5.6	22.3
D·A·C	see	1.5.6	1.22.3
D·A·E	see	3.5.6	21.22.3
D·B·C	see	1.4.6	1.211.3
D·E	see	3.6	21.3
D·E·C	see	1.3.6	1.21.3
D·F·C	see	1.2.6	1.2.3
E·C	see	1.3	1.21
F·C	see	1.2	1.2

STRUCTURAL CRITERIA

The preceding analysis suggests a number of criteria which may be used in describing the structure of retrieval systems, as follows.

1. The lattice relations incorporated, whether a search for term T can also retrieve as "relevant" terms *(a)* included in T, *(b)* including T, *(c)* coordinate with T. The traditional hierarchical classification incorporates all these relations; a system using unique descriptors incorporates none of them; a cross-referenced alphabetical index may incorporate them partially. Measures of the "dimensionality" of a system are needed, the average length of inclusion chains, the size of coordinate arrays.
2. The type of subject indexed, simple or compound. Simple (single-term) subjects are typical of the purely enumerative classification, but most retrieval systems index compound subjects. The "connectivity," maximum and average numbers of terms in a compound, would be characteristic of a system. The capacity to form interlocked compounds in subjects is also a characteristic feature.
3. The form of entry, by term or by item. Typical item entry systems are library catalogues, and term entry systems have been devised by Batten, Taube, and others.
4. The relationship of entries to each other, fixed or free. The typical fixed entry system is again a library catalogue, while punched card systems usually have entries free.
5. The arrangement of terms in term entries, fixed, free, or in fields. A composite subject heading is typically fixed; superimposed coding on cards has terms free; other punched cards use fields.
6. The symbolism used for terms, verbal, random code, ordinal class, or hierarchical class symbols.
7. The presence or absence of a chain index, a feature which arises from characteristic (5).
8. The type of search operations which can be conducted, logical sums, products, differences, and combinations of these. This feature has not been explicitly considered here, as it is now familiar in retrieval theory.

By permuting these eight sets of characteristics, a large variety of different retrieval systems can be designed, even though not all the characteristics are independent. The list of criteria is of course not intended to be exhaustive, it includes only abstract structural features of the system, and does not necessarily cover all the possibilities even in this sphere. Once the physical mechanisms and operation of a retrieval system are included, a whole host of other criteria arises, and many of these have been listed in the "Proposed definitive statement of the scope of area 4." These may be excluded from consideration in a study of general structure.

OPERATIONAL EFFICIENCY

The formulation of criteria by which to describe the structure of a retrieval system is not an end in itself. It is only a necessary preliminary to discovering which structural features lead to the highest operational efficiency of the system. The next step is to analyse the operations associated with the various structural features. The operations are of several kinds. First to be considered are operations needed to design the system:

1. Constructing a lattice of units of information. This is a *semantic* problem and involves the formulation of postulates as to the kinds of terms and relations to be used, and the subject analysis of the chosen field in the light of these principles.
2. Constructing schedules (L_1 to L_5) by appropriate transformations of the lattice, and the required coding. Next there are operations needed to construct the index:
3. Analysing the subject matter of each item into appropriate lattice units, terms.
4. Consulting schedules (L_1, L_2, L_5) to ascertain coding.
5. Inscribing index entries with term symbols and item specifications.
6. Filing index entries.
7. Constructing chain index entries.

Lastly there are operations needed in using the system:

8. Naming the subject sought in appropriate terms.
9. Consulting schedules (L_2 and L_5) to ascertain coding.
10. Consulting chain index.
11. Inscribing search entries.
12. Matching search entries with index entries.
13. Moving to related index entries.
14. Inspecting matched and related index entries and recording inscribed item specifications.

Of these fourteen operations (which are, of course, subject to more detailed breakdown), numbers 1, 2, 3, and 8 must be performed by human agents. For the same indexed material and search questions, they will vary almost entirely in accordance with variations in the lattice. For all systems using the same lattice, these operations will be almost identical. Operations 12, 13, and 14 are purely *clerical* operations within the indexing system, and will vary widely according to the physical mechanisms involved. The remaining operations, 4, 5, 6, 7, 9, 10, and 11, are also clerical. As listed above, a human clerk is assumed to be consulting, inscribing, filing and chain indexing, but all these operations could be carried out by a suitably instructed machine, and would then vary, not only in accordance with the physical nature of the index entries but also in accordance with their own mechanical nature.

The relative efficiency of different ways of carrying out these operations, therefore, cannot be discussed without reference to *(a)* the semantic problems of lattice construction, and *(b)* the physical mechanisms or clerical hardware. Neither of these subjects can be considered in this paper. Nevertheless, it is useful to conclude this study of the general structure of retrieval systems by examining some meanings of the word “efficiency.”

Two aspects of efficiency must be distinguished. There is first the *retrieval efficiency,* i.e., the degree to which a system actually retrieves from its store those item specifications which are in fact relevant to the search question. There is secondly the *economic efficiency,* i.e., the cost in labour, materials, time, and money of achieving a given retrieval efficiency. The economic efficiency will be a function of all the factors so far discussed: the semantic nature of the lattice, the structural criteria of the system, and the clerical hardware. It can be assessed by economic studies of existing complete retrieval systems, and by partial studies of particular, relatively self-contained operations within such systems.

The retrieval efficiency is dependent on fewer factors. The different forms of clerical hardware are only different ways of physically embodying certain structural characteristics. These in turn are alternative ways of transforming a given matrix and lattice. In principle, every retrieval system based upon a given lattice should have the same retrieval efficiency. Any variations should be introduced only as *engineering noise,* e.g., mistakes in coding by human agents, errors in machine operation, or “false sorts” deliberately allowed in the design of the system.

Perry and his associates have suggested several measures of retrieval efficiency, e.g., the *noise, factor* (the fraction of retrieved items not relevant to the subject sought) and the *omission factor* (the fraction of actually relevant items in the system which are not retrieved). These factors could be determined by examining, respectively, the retrieved items and the whole collection of items indexed. It is necessary to consider the relation of these factors to the meaning of relevance identified earlier.

If items are retrieved whose entries actually bear the sought combination of terms, but which prove to be irrelevant, then we are dealing only with engineering noise, a mistake has been made either by the indexer or the searcher. True *semantic noise* occurs only when an entry bearing terms *related* to the sought terms is found not to be relevant. This may mean that in designing the system and constructing the lattice, the *relevance limits* have been made too wide, too great a selection of inclusive, including, or coordinate subjects has been linked to the sought subject. Or it may mean that in this particular search, the relevance limits have been made too wide, though in other searches they may not be wide enough, in which cases there will be significant omission factors.

The implications of tests of the retrieval efficiency of a system are therefore clear. A persistently and inconveniently high noise factor implies that the relevance limits should be narrowed; a persistently high omission factor, that they should be widened. In each case, detailed analysis is necessary to decide the inclusion or coordinate relations which need to be amended. If some tests on a system give a high noise factor, and others a high omission factor, the implication is that the system should be redesigned to permit variations in the relevance limits according to the search required.

ACKNOWLEDGMENTS

Since this paper has been written in very general terms, it has not been found necessary to refer to many other writers. Nevertheless, much is clearly owed to those workers in information retrieval whose work I have tried to generalise.

It is no accident that of the few names cited one should be R.A.Fairthorne, from whom both the lattice and matrix concepts have been drawn. An exposition of a retrieval system as a lattice is found, for example, in his paper on "The patterns of retrieval," *American Documentation, 7* (1956), 65–70, and a brief account of a talk on the matrix in *Classification Research Group Bulletin,* 1957, no. 3.

The reference for measures of retrieval efficiency is to J.W.Perry, A Kent, and M.M.Berry, *Machine Literature Searching* (Interscience, 1956), which also contains an extended account of searching operations in terms of products, sums, and differences. The reference to M.R.Hyslop is to her chapter in J. H.Shera, A.Kent, and J.W.Perry, *Documentation in action* (Reinhold, 1956).

The concept of interlocking is implicit in the indexing system of J.E.L. Farradane, *J. Documentation, 6* (1950), 83–99, and explicit in the paper of C.N. Mooers, "Information retrieval on structured content," in *Information theory* (C.Cherry, editor, Butterworths, 1956). It is also implicit in the "role indicators" of Perry *(loc. cit.)* and the "interfixes" of D.D.Andrews and S.M.Newman, *Storage and retrieval of contents of technical literature, nonchemical information* (U.S. Patent Office, Research & Development Report, 15 May 1956).

The problems of term symbolism have been dealt with by many writers, but

the work of E.J.Coates is of particular value: see *Classification Research Group Bulletin,* 1956, no. 1; 1957, no. 2. S.R.Ranganathan has written copiously on symbolism, citation order of interlocked terms, chain indexing, and other problems. Four of his books may be cited: *Classification, coding and machinery for search* (Unesco, 1950), *Classification and communication* (Delhi University, 1951), *Philosophy of library classification* (Copenhagen, Munksgaard, 1951), and *Prolegomena to library classification* (London, Library Association, 1957).

Reference may also be made to B.C.Vickery, *Classification and Indexing in Science* (Butterworths, London, 1958) and to papers cited therein; and to the *Classification Research Group* memorandum, "The need for a faceted classification as the basis of all methods of information retrieval," reprinted in *Proceedings of International Study Conference on Classification for Information Retrieval* (Aslib, 1957).

Note: Term entries have also been called "aspect" or "feature" entries. "Unit cards" are term entries which are further broken down so that each bears only a single item specification.

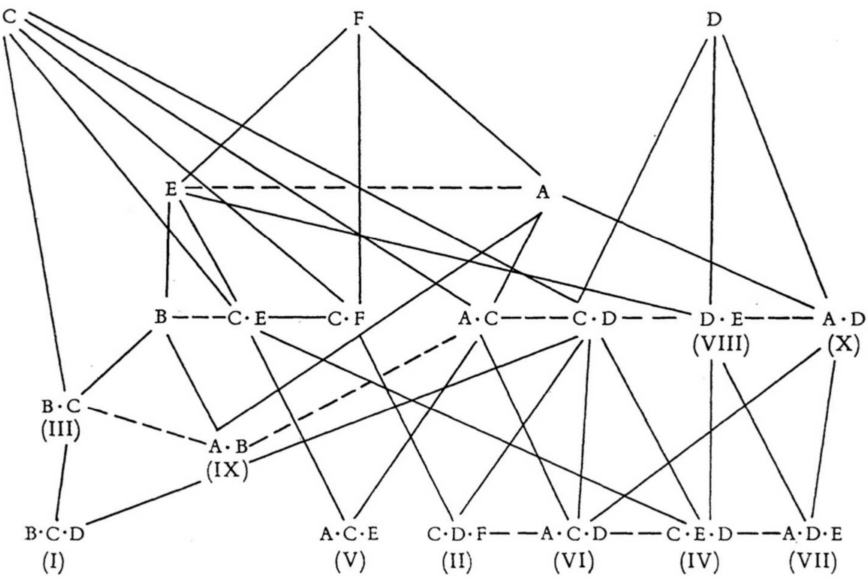


FIGURE 1. An information lattice showing inclusion (F—E), coordinate (E – – – A) and interlocking (B · C) relations between subjects.

The Descriptive Continuum: A “Generalized” Theory of Indexing

FREDERICK JONKER

ABSTRACT. This article first notes the importance of the indexing problem in mechanized information retrieval: not only is the degree to which information can be retrieved entirely determined by the adequacy of the system of indexing, but the total cost of the mechanized information control system is to a large extent likewise determined by the indexing system.

In view of this overriding importance of indexing systems, the paper points out the desirability of general criteria common to all indexing systems, which can be applied to any possible indexing system in evaluating its suitability to a particular information control problem.

In an attempt to find such criteria common to all indexing systems, and establish a “generalized theory” of indexing, the history of information control is traced from the earliest classification systems through subject heading systems to the latest developments such as Uniterm indexing.

The various causes underlying the evolution of the art are likewise analyzed. In the light of this analysis the basic cause of an information retrieval problem and the degree of severity of the problem lie in the degree of *diffuseness* of information. Diffuseness is the degree to which required information is contained within other, only indirectly related, information items and subjects. The greater the degree of diffuseness, or unrelatedness, the more severe is the problem of information control.

In the historical analysis it is shown that the increasing diffuseness of information forced the art away from classification systems to subject heading systems, and from these to simple keyword indexing. It is shown that there is a gradual transition from one system to the other and that no sharply delineated basic differences exist between any of these systems.

The generalized theory of indexing postulated in this article therefore looks upon all indexing systems as a continuum, the *descriptive continuum*. The main parameter of this continuum is the average length of the “entries” or “headings” used. At one end of the continuum or “spectrum” is keyword indexing; subject heading indexing is somewhere in the middle, while hierarchic classifications are at the other extreme. The average length of the

---

FREDERICK JONKER Documentation, Incorporated, Washington, D.C.

The work reported here was done under the sponsorship of the Directorate of Advanced Studies, Air Force Office of Scientific Research, Contract No. AF 49(638) 91.

headings or descriptive terms used determines the position in the continuum.

Throughout the continuum, all other parameters behave as functions of the average term-length. Some of these parameters are:

Potential depth of indexing

Permutability of indexing criteria

Degree of hierarchial definition of indexing

Potential need for a coordinating mechanism

Retrieval noise

Size of the access apparatus

False coordinations

Capacity for handling semantic indeterminacy

These parameters are discussed and explained. They are believed to contain all the considerations basic to the indexing problem.

The theory indicates that once the main parameter, average term length, is determined, all other properties of the indexing system are fixed. For every information collection there is an "optimum" position in the continuum, according to which the collection should be organized. This optimum position is determined by the diffuseness of the information in that particular field.

In any mechanized system of information storage and retrieval there are three distinctly separate problems:

1. The *indexing* problem. The problem of assigning to a certain item of information a number of "terms" characterizing this item of information. The word "term" is used here in its broadest sense and comprises any form of "class," "subclass," "subject heading," single words or combinations of words. The indexing problem is common to manual as well as mechanized information control systems. In non-mechanized information control systems the following two problems do not enter.
2. The *coding* problem. The problem of converting terms into a code so that they can be stored in a machine.
3. The *machine systems* problem. The problem of selecting a machine system which can store and manipulate the coded information.

This paper is concerned with indexing. The determination of the particular form of coding and machine system which will best suit the requirements of any particular problem at hand is basically a matter of engineering expediency. Ultimately it reduces to problems of cost. Almost *any* compatible combination of coding and machine system can in principle perform *any* required form of information search. However, some machines may be better suited to a particular information control problem than others. The "best" system is simply that with the lowest total cost for the required information entry and search operations.

All problems in information storage and retrieval can ultimately be reduced to one single factor, cost. All "improvements" in this field, and many other fields, basically aim at one thing and that is lowering of the total cost of a system able to meet a certain specified performance.

Mechanization would seem to be the answer to this problem. But in general, the cost of the intellectual work of indexing is roughly on the same order of magnitude as the combined cost of the clerical work of entering the data in a machine and the cost of the information control machine itself. Therefore mechanization can be only a partial answer. The other part to the answer must be found in more economical systems of indexing which provide greater depth of indexing at lower cost. It is this consideration that gives the problem of indexing its tremendous importance.

The various systems of indexing in use today confront us with a picture of conflicting claims. Rather than single out and discuss their differences, we will here attempt the opposite and look for elements which these apparently conflicting systems have in common and attempt to establish one single "unified" or "general" theory of indexing.

The suspicion that such a general theory must exist seems to find confirmation in the pattern of evolution of many of the sciences, particularly the physical sciences. New fields of science usually start with the observation and measurement of a number of freshly discovered phenomena, which are vaguely felt to be related. As the new science progresses, closer and closer relationships between these phenomena are discovered. Finally a generalized theory and generalized equations become possible. Many sciences which do not lend themselves to mathematical quantification exhibit basically the same pattern of development. The evolution of a science always seems to progress from the understanding of a large number of relationships, each linking a limited number of phenomena to one single relationship relating all observed phenomena.

The inescapable conclusion seems to be that no true understanding of existing indexing systems and problems seems possible, unless all systems can be seen in the light of more general common concepts, linking all these systems together into a single "closed" system.

One fruitful way to investigate the elements in common among the various systems of indexing is to make a brief study of their origin in history, of their evolution in time. Below we attempt to trace not only the actual events but also their underlying causes.

HISTORY AND ANALYSIS OF THE INFORMATION CONTROL PROBLEM

Until a few centuries ago information control problems, as we know them today, were entirely unknown. The totality of recorded knowledge, at the time of the early book libraries, was still extremely small. Few books were written and each could easily be classified in one or more of the few then recognized sciences and fields of human endeavor.

Although the origin of classification systems goes far back in history, at no time was much significance attached to classifications until the 18th century "enlightenment" brought about a greatly increased knowledge of the physical world. Most of this new knowledge, to a large extent of taxonomic nature, lent itself readily to classification. Cataloging and classifying became at that time one of the most important goals of science. Classification systems gained an enormous importance. The hope that all knowledge necessary for the understanding of the physical world would soon be complete and fit into a "natural" and "universal" classification system became one of the fondest dreams of the age.

The importance of classification systems was not, however, based on considerations of information storage and retrieval. Such problems at that time hardly existed. It was based solely on the urge to see order in knowledge and to organize all knowledge into one "closed" system. Although subsequent developments relegated classification systems to a far less important position, vestiges of the enormous prestige gained at that time persist today.

The industrial revolution created a new body of knowledge *entirely different in nature and much larger,* by many orders of magnitude, than all preceding knowledge. Moreover the center of gravity of human knowledge shifted from the natural sciences to the field of technological and supporting sciences; to the knowledge of how to build railroads, ships, airplanes, and powerplants; how to cultivate the soil, cure diseases, etc. Not only did the nature and size of our store of knowledge change radically, but the need to consult the store of information rose rapidly.

It was this *technological* information which for the first time created a problem of information storage and retrieval. Relatively few problems are encountered in "consolidated," well-organized information like that contained in encyclopedias, handbooks, and textbooks, where information pertaining to a certain subject has been collected from a great many different sources under one single heading. Only when information is scattered through a great many different articles and reports and when each of these "items of information" contains bits of intelligence related to, or of potential importance to a number of different subjects, does an information control problem arise. It is this *diffuseness* of technological information which causes and defines the information control problem as it faces us today.

Diffuseness of information implies that items of information may be of interest in a great many problems, uses, and fields of pure and applied science.

It implies that these items should be indexed from each viewpoint of potential importance. The diffuseness of an item of information therefore is measured by the number of potential indexing viewpoints or criteria.

By far the greatest proportion of our present store of knowledge which we need to consult constantly and rapidly is of a highly diffuse nature. This high degree of diffuseness is not something accidental. It is inherent in the very nature of our technological civilization, in the fact that every achievement or finding in any particular field may be of enormous potential interest in almost any other field. It is this potential cross-fertilization which is responsible for the rapid acceleration of progress, which, measured by the time scale of history, almost assumes the proportions of an explosion. Here indexing theory touches the mechanism of progress.

The degree of diffuseness of information is the heart of the indexing problem and the main parameter in the generalized theory of indexing. If an information collection is assembled for very special purposes only, the collection need of course not be indexed to its fullest depth, that is, the indexing will cover only a fraction of the full potential implications of the information. Generally, however, it is desired to index so as to encompass the full potential. In that case the "indexing depth" should cover all foreseeable implications. It should equal the diffuseness of the information.

In assembling the large libraries of the last century so as to contain all the new knowledge, the need for a system of information storage and retrieval was immediately felt. The well-established classification concept was of course the first device these libraries turned to. But it was soon discovered that in the new fields of technology and supporting sciences no classification, no matter how elaborate, could satisfy even the most straightforward demands.

To understand the reason for this sudden breakdown of classification systems in their application to technological information, it is necessary to analyze briefly the problem of "class-inclusion" or, more generally, the problem of "subordination." Although most severely felt in classification systems, it is basic to any form of indexing.

The simplest form of indexing is encountered when an object is indexed according to only *one particular criterion* or from *one particular point of view*. In such a case it is frequently possible to set up mutually exclusive classes. Even when no fully mutually exclusive classes are possible, the degree of overlap of these various classes will in general be small, and serious problems are not encountered.

For example, in the Patent Office where the main problem has centered around *structural* differences and similarities, it has been possible to maintain until today a classification according to mechanical structure. This has not been accomplished without very serious problems, as even the analysis of mechanical structure is subject to differences of point of view. Mutually exclusive classes exist only to a limited degree.

On the other hand, most of our present information collections consist of highly diffuse information which requires indexing from many viewpoints, such as: *design components, design principles, possible uses and applications, materials used, physical principles involved, sciences involved, production methods involved,* etc. Each viewpoint requires an entirely different form of subordination. The technological and supporting sciences have a practically *unlimited number of index criteria* or viewpoints from which they can be indexed. Some of them may seem extremely trivial. But this apparent triviality disappears in the realization that even the most specialized viewpoints often represent billion-dollar industries to which such parochial viewpoints are vital. As the technological evolution progresses, fresh viewpoints constantly emerge, and older viewpoints change their relative importance or disappear.

It has been strongly felt in the past and is still maintained today that although many varying "special" indexing viewpoints of secondary importance do exist there must be one "absolute" or general indexing criterion which can serve every purpose equally well. "Structure" is often quoted as an example of such an absolute criterion. In the past century, mainly a time of mechanical technology and mechanical (Newtonian) world concept, "structure" has seemed to overshadow all other considerations. But the recent shift of emphasis in physics has shown the relativity rather than absoluteness of this Newtonian viewpoint. There are also in many fields of activity special indexing criteria which seem of preponderant importance, mostly because they were historically the first to emerge and the prevalent nomenclature is still based on them. However, for the purposes of the technological sciences, which present an ever widening and changing spectrum of indexing criteria, no basically absolute criteria nor absolute classifications can be recognized.

The implications of a high degree of diffuseness of information can be illustrated by a very simple classification problem: living beings can be indexed or classified according to a large number of criteria, for example: *(A)* skeleton structure, *(B)* reproductive system, *(C)* circulatory system, *(D)* digestive system, *(E)* skin structure, *(F)* habitat, *(G)* usefulness to human society. Each criterion or "viewpoint" can form the basis of a "classification." Figures 1 and 2 show schematically a classification based on criteria *A* and *B,* respectively. This form of classification can be termed "single-criterion classification," The classes A_1, A_2, A_3, etc., or the sub-classes A_{11}, A_{12}, A_{13}, etc., could of course be arranged alphabetically. However, there usually is a more meaningful order inherent in the indexing criterion. This order can, for example, be based upon a size relationship or a relationship of time or space, on some causal relationship, or on a combination of relationships. For example, in drawing up a classification of internal combustion engines upon the criterion of "structure," we can start at the input side of the engine and enumerate all the parts as they contribute and relate to the power generating process. We would in that case start

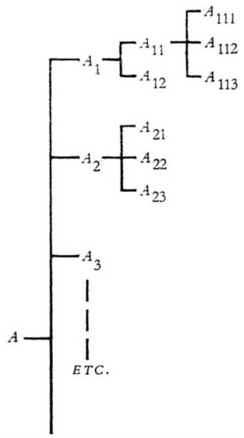


FIGURE 1. Single criterion classification. Classification based on indexing criterion *A,* skeleton structure.

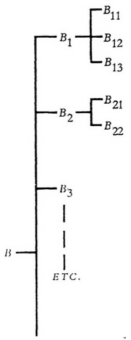


FIGURE 2. Single criterion classification. Classification based on indexing criterion *B,* reproductive system.

with the carburetor, and go through the valves into the cylinders; from there on through the pistons, connecting rods, etc., to the power output shaft. This would be a subordination based on a causal principle. To cover all parts of an engine completely, a combination of causal and spatial relationships would be required.

Each of the seven single-criterion classifications of living beings in the example given above, are incompatible and entirely different from the others. Each of these can serve only one particular field of interest or type of user. A "universal" classification, of interest to all possible users, must therefore integrate all these points of view through indexing by all seven criteria. Figure 3 shows schematically a classification based on a combination of only two different indexing criteria. It is seen that the larger the number of indexing criteria, the more complicated such a classification becomes. Although in practice only a small part of all possible combinations will actually be used, the classification system attempts to provide sufficient classes for all anticipated needs. The task is a staggering one. The added task of keeping it up to date by providing for indexing by newer or changed criteria is still more staggering and the cost prohibitive. In practice more than two indexing criteria are rarely used.

Yet the main problem in drawing up a classification system still lies elsewhere, namely, in the decision as to which criterion should be elevated to constitute the main class, and which criteria should constitute respectively the *sub*-classes and *sub-sub*-classes.

If, in the example, it should be decided to use criterion *A,* "skeleton structure" to draw up the main classes, and criteria *B,* "reproductive system" and *C,* "circulatory system," etc., for the sub-classes and sub-sub-classes respectively, this classification would be most useful from the point of view of bone specialists. However, a user interested in criterion *G,* "usefulness to human society," would have to consult all sub-sub-classes to find all information in the system of potential interest to him.

Where a very specific purpose is concerned, this form of classification can provide quite a useful tool. For example, for a research laboratory interested in data on stress failures of engine parts, a classification of engines based on engine structural components and parts as a main indexing criterion and the various forms of load and stresses responsible for failures as sub-criteria would serve the purpose admirably.

The above form of classification can be termed an "unpermuted multiple-criterion classification." In such unpermuted classifications, each item of information need be classified in only one way. No cross-reference structure is required. If an article contains more than one item of information, for instance, information on heat stress in turbine wheels and on torsional vibration stress in

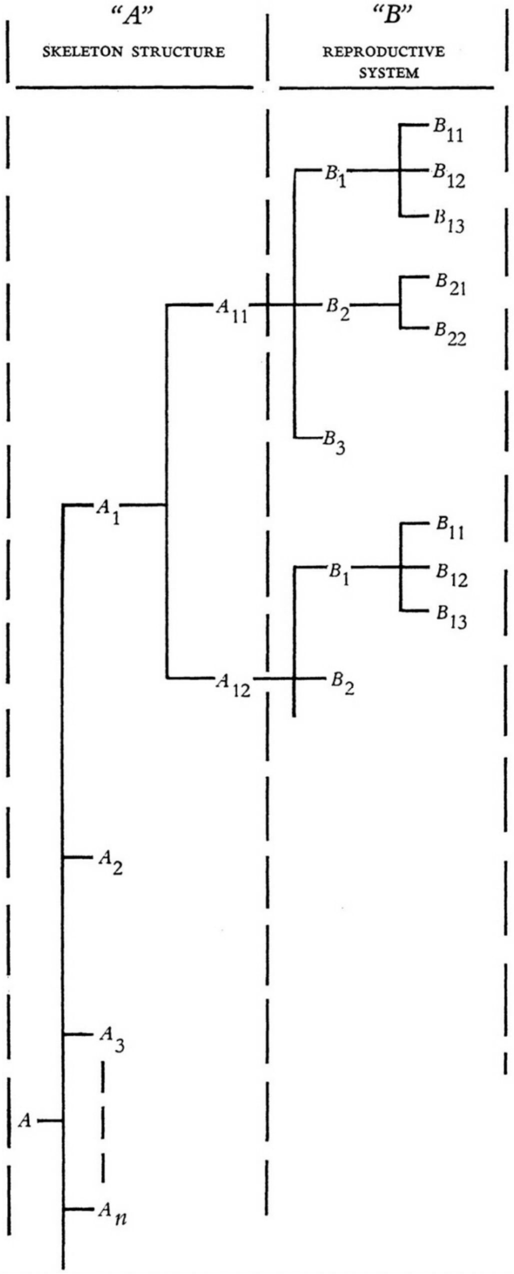


FIGURE 3. Unpermuted multiple-criterion classification. Classification based on criterion *A* as main criterion and criterion *B* as a sub-criterion.

crankshafts, each of these items of information will of course have to be indexed distinctly and separately from the other. However, no cross-reference structure is yet called for.

The unpermuted multiple-criterion classification is the most frequently used form. It is generally considered to be the "purest" form of classification. For the purposes of the generalized theory it will be considered the definition and general model of a classification. The reasons for this will become apparent below.

Unpermuted multiple-criterion classifications can meet only special purposes, those in which the subordination of indexing criteria lends itself to the requirements of the special purpose, as in the example of engine failure data. A "universal" information retrieval system to meet *all possible requirements* would, therefore, have to consist of a larger number of different classifications, each with a different hierarchial subordination or "permutation" of the indexing criteria. In size and cost this would be a truly astronomical proposition.

Instead of keeping the different classifications separate, it would seem natural to integrate or merge them into one "universal" system. But when attempt is made to effect such a merger it is found that there simply are no relationships between the classes (other than that of their symbols) upon which their order can be based. In other words, there is no real practical alternative to merging the main classes A_1, A_2, $A_3 \cdots$ and B_1, B_2, $B_3 \cdots$ of all these classifications alphabetically according to the first letter of main classes.

However, in doing this the concept of meaningful hierarchial relationships is abandoned. Even small amounts of permutation virtually force this abandonment. Therefore permuted classifications can no longer be considered "true" classifications.

Far more serious, in achieving the final merging the very purpose and advantage of a classification, namely, *ease of searching,* is defeated. Not only has the size become prohibitive, but the searching ease is no better than in a conventional subject heading system. Such a "classification" is *exactly identical* to a theoretically complete, fully permuted subject heading system. In practice, a permuted classification can never hope to be complete, and is extremely fragmentary at best. Permuting a classification requires that each item of information be classified in as many ways as the number of permutations of the indexing criteria on which the classification is based. It also requires a corresponding cross-reference structure.

Summing up, it can be said that classification systems have been found practical only under two conditions: *(a)* where there is an *unambiguous* and *unchanging* subordination of the indexing criteria as well as of the retrieval criteria; *(b)* where the indexing criteria are *fixed and very limited in number*.

In other words, hierarchic classifications are practical only where information of a low degree of diffuseness is concerned and the system is intended only for one particular special use.

The difficulties encountered in classifications have intensified as the degree of diffuseness of the information has increased. As a result libraries have tended more and more toward alphabetic systems, although there remains in Europe a strong core of support for classification in general and the U.D.C. in particular.

This transition was a gradual process leading to a continuing replacement of hierarchial classification by alphabetical subject heading systems. Many existing classifications actually are mixed forms which show many characteristics of the subject heading list. The *alphabetic subject heading* list relieves the system of the embarrassing necessity to decide which criteria to subordinate to other criteria. It provides a less bulky and less expensive apparatus that is easier to maintain and has sufficient flexibility for incorporating new indexing criteria.

With the emergence of special libraries serving industry and Government and with the related development of technical information and research data centers, alphabetic subject headings were in turn challenged by "descriptors," "Uniterms," or keyword indexing. Where the need was felt for the use of terms consisting of a combination of two or more single words, some form of mechanical "combinatory," "coordination," "collation," or "coincidence" device was used. These devices made it possible to search by all possible combinations and permutations of terms, without the need of incorporating these combinations and permutations in the subject heading list.

Before analyzing the merits of keyword indexing, the two main components of an information storage and retrieval system will be briefly discussed. These components are:

1. The *information store* or the memory. In a conventional system this would correspond to the card catalog consisting of the filing cabinets with the catalog cards.
2. The *access* apparatus, which provides for proper "access" to the memory. In a conventional system this would correspond to the printed subject heading list or printed classification schedules, usually in book form.

Although these two components can be physically combined, it is always possible in any information storage and retrieval system to distinguish between their two functions.

Keyword indexing as formalized, for example, in the Uniterm system, shows the most spectacular advantages in setting up new libraries. No expensive pre-made subject heading list nor classification schedule is necessary nor is it hard to acquire familiarity with the access apparatus. Keywords or ideas taken from or suggested by the text are used as index terms. After the collection has reached a certain size, the vocabulary of terms will automatically retard its growth and will stabilize itself. The only "maintenance" required is the provision of a cross-reference structure to take care of synonyms, etc. In its simplest form keyword indexing provides the greatest possible retrieval power, as any single term or any combination of index terms can be used to retrieve an item of information.

The cost of such systems usually is only a small fraction of the cost of a conventional system.

If properly indexed, an item of information is indexed by any keyword that is or could possibly become of importance to any potential user of the item of information. Therefore it is also automatically indexed by any or all possible combinations of these terms. The combinations normally include all subject headings or classes and subclasses by which this item of information would have been indexed, had a subject heading list or a classification system been used.

These advantages are, however, bought at a price. This price is the acceptance of a rudimentary "access apparatus," usually consisting of no more than an alphabetic listing of the complete vocabulary of index terms in use in the system. Such a listing cannot directly tell the user, with a generally desired degree of assurance, what information is contained in the system. For example, a vocabulary of index terms listing the term "water" as well as the term "pump" does not necessarily imply that the system contains any information on water pumps, nor does the presence of the term "jet" as well as the term "engine" necessarily imply the presence of information on jet engines. In other words, there is no way of ascertaining a priori whether the combined terms "water pump" and "jet engines" are "true" or "false" combinations, until the search process is completed.

Another disadvantage may be the lack of *hierarchial definition* within each indexing criterion. Indexing tends to be performed at the level at which the concept happens first to be referred to, in the text. For instance, an Army General indexed by the criterion of military rank is simply indexed by the term "generals" instead of "field officers," or "officers" or "soldiers," or any combination of both. A broad search for everything on "soldiers" may therefore require a number of individual searches by each of the above terms. For these purposes a cross-reference structure has to be maintained in the access apparatus. As a result of these weaknesses of the access apparatus an exhaustive search by a combination of terms may require a number of individual searches by various combinations of various terms. Therefore the low indexing cost is obtained at the price of somewhat higher search costs. However, since in most information collections the data input load is far greater than the output or search load, this theoretical disadvantage is of little practical significance.

It is possible to alleviate this weakness of the retrieval mechanism by using longer index terms. Instead of the elementary terms "water," "pump," "jet," and "engine," combined terms like "water pump" and "jet engine" can be used. A vocabulary of index terms will in this manner become more suggestive of the contents of the information system and fewer individual searches will be required to obtain full search results.

Lengthening the average index term produces changes in two other parameters. The first change is a reduction of retrieval power. In using the combined terms "water pump" and "jet engine," we can no longer retrieve complete information by the broader terms "pumps" or "engines." Also, if the item of information likewise dealt with water jets but was *not* indexed by this term, as water jets seemed at *that* time of no interest to the users of the system, the possibility of retrieving by the term water jets at a later time is lost. Indexing by the elementary terms "water," "pump," "jet," and "engine" would have preserved this possibility.

The other parameter change accompanying a lengthening of the index terms is a reduction in the false combinations of terms. With the true combination "water jets," the false combinations of, for example, "water engine" and "jet pump" have likewise disappeared.

If this process of lengthening of the index terms is continued, the combination of terms may eventually attain the length of full-fledged subject headings. The need for retrieval by a combination of two or more of these single terms will have been greatly reduced, while the alphabetic listing of the vocabulary of index terms will have attained the structure of a subject heading system.

At the end of this historical analysis of the indexing art a few words on "meaning" seem desirable. It is urged by some, that the information storage and retrieval art devise means of indexing by "meaning" or "true classes" instead of just "words." The substance of this problem can be stated as follows: "meaning" is only arrived at by a common agreement on what certain terms will be considered to mean or what terms will be used to describe certain notions. This is in essence standardization of language. Such standardization is from time to time undertaken by the various learned bodies and engineering societies. Without some form of standardization, whether organized or spontaneous, no intelligent communication of any kind is possible, and any form of communication should use standardized language as far as available.

Certainly the indexing art should use standardized language to the fullest possible extent. However, standardization of definitions and the process of indexing are entirely different. Standardization of meaning should preferably be undertaken by or sponsored by the societies representing the various fields of endeavor, so that this standardized meaning may become universally acceptable. The information control art should, where possible, promote standardization of meaning. It should, nevertheless, be realized that the information storage and retrieval art by its very nature finds its greatest use at the frontiers of progress in fields of research with new and not yet consolidated concepts. This is precisely the area where new words and concepts are constantly being formed.

Standardization of meaning will always be far behind the emergence of these new concepts, as a period of crystallization and stabilization is always required before standardization can be successfully undertaken. Therefore, any indexing system used on "frontier" information should be able to function even if no sharply defined meanings have yet been established. This would seem to be a *conditio sine qua non* for indexing advanced research data. Keyword indexing would seem to meet this requirement. In the absence of single well-defined terms, a number of possible alternative terms can be used. On the other hand, a classification having a hierarchial subordination requires well-defined terms in order to determine this subordination. Hierarchial classifications therefore seem unable to function in areas where semantic indeterminacy still prevails.

Summing up, indexing by "meaning" is not a difficult problem where standardized terminology is available. The *real* problem and challenge is indexing in areas where no such standardized terminology is yet available, and in making information retrieval systems work despite the absence of standardized language.

GENERALIZED THEORY OF INDEXING

DEFINITION OF THE CONTINUUM

In the preceding sections, the history and evolution of information retrieval problems has been traced. During this evolution the entire range of possible systems from classification to indexing by the smallest possible terms has been traversed. The entire spectrum of possible indexing systems having thus been laid bare, the formulation of a generalized theory now becomes possible.

The evolution has been a gradual process during which one system automatically developed into another. But it must be recognized as an irreversible, unavoidable process which took place under the pressure of changing conditions and needs.

In the light of this analysis the various known indexing systems appear to form a *continuum* rather than contrasting systems. As the function of this continuum is the "description" of information contained in the collection by means of "terms" suited to adequate information storage and retrieval, this continuum might further be referred to as the "descriptive continuum."

The descriptive continuum is schematically illustrated in Fig. 4, *a, b,* and *c*. The parameter which determines *position* in the continuum is the *average length of index terms,* measured simply as the average number of letters per term throughout the information collection. This has been selected as the main parameter of the general theory of indexing, defining the continuum (Fig. 4*c*). Relative to this parameter, the various types of indexing find their respective positions in the continuum roughly as shown in Fig. 4*b*. Depending upon the width or narrowness with which we define each of these systems, the areas covered by their designations either overlap to some extent, or leave a certain amount of gap. However, no matter how narrowly we define the areas of these systems, the entire spectrum is in practice actually used.

One end of the continuum represents indexing exclusively by the *smallest* single words available in the language. That is, it uses the smallest "units" available. Such a system will therefore use the largest possible number of terms per item indexed. At this position in the continuum we would, for example, index an article on the "erosion of turbine buckets in bunker-C oil burning jet engines" by the individual single terms "erosion," "turbine," "bucket," "bunker-C," "oil," "jet," and "engines." This is the area of keyword indexing in its purest form.

On moving away from this end of the descriptive continuum, the average length of the index terms increases. In the preceding example individual terms like "jet," "engine," "turbine," and "bucket" would gradually be replaced by combined terms like "jet engines," and "turbine buckets." This takes us gradually out of the area of "keyword" indexing into what starts to resemble a conventional subject heading list. As the average length of the terms continues to increase and we begin to use terms like "jet engine, turbine bucket" or "turbine bucket, jet engine," and "erosion, jet engines," we have reached the area of the *subject heading list*.

The other extreme of the continuum represents the area of hierarchic *classification,* where we index by the longest possible terms and only one term is required to index an item of information. At this position in the continuum, an article on the "erosion of turbine blades in bunker-C oil burning jet engines" would be indexed by classes, sub-classes, sub-sub-classes, etc., for example as follows: "internal combustion engines, steady-state engines, gas turbines, turbine blades; liquid fuel, mineral fuel, bunker-C oil; wear, erosion." For the purposes of this general theory, the above classification must be considered one

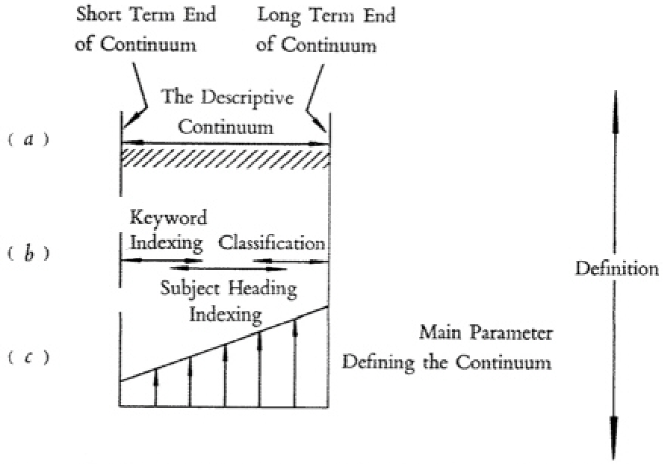


FIGURE 4. (*a*, *b*, *c*) The descriptive continuum. (*c*) Average length of index terms.

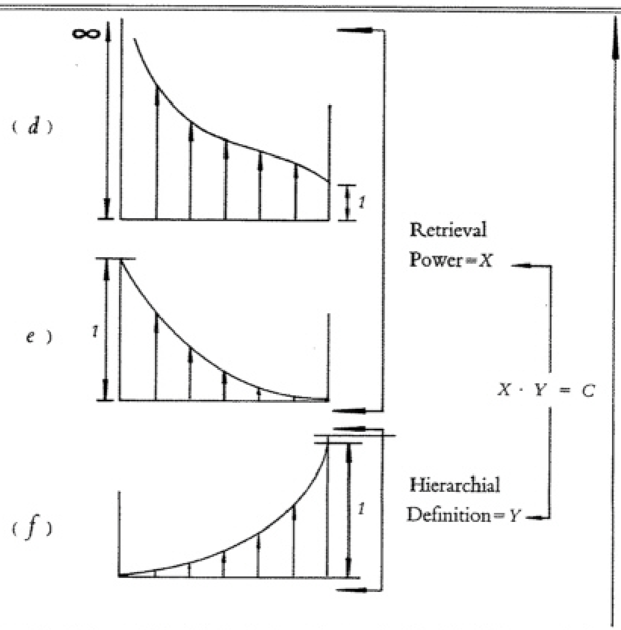


FIGURE 4. (*d*) Potential depth of indexing and retrieval; allowable diffuseness of information; number of terms per item of information. (*e*) Permutability of retrieval criteria. (*f*) Degree of hierarchial definition of indexing.

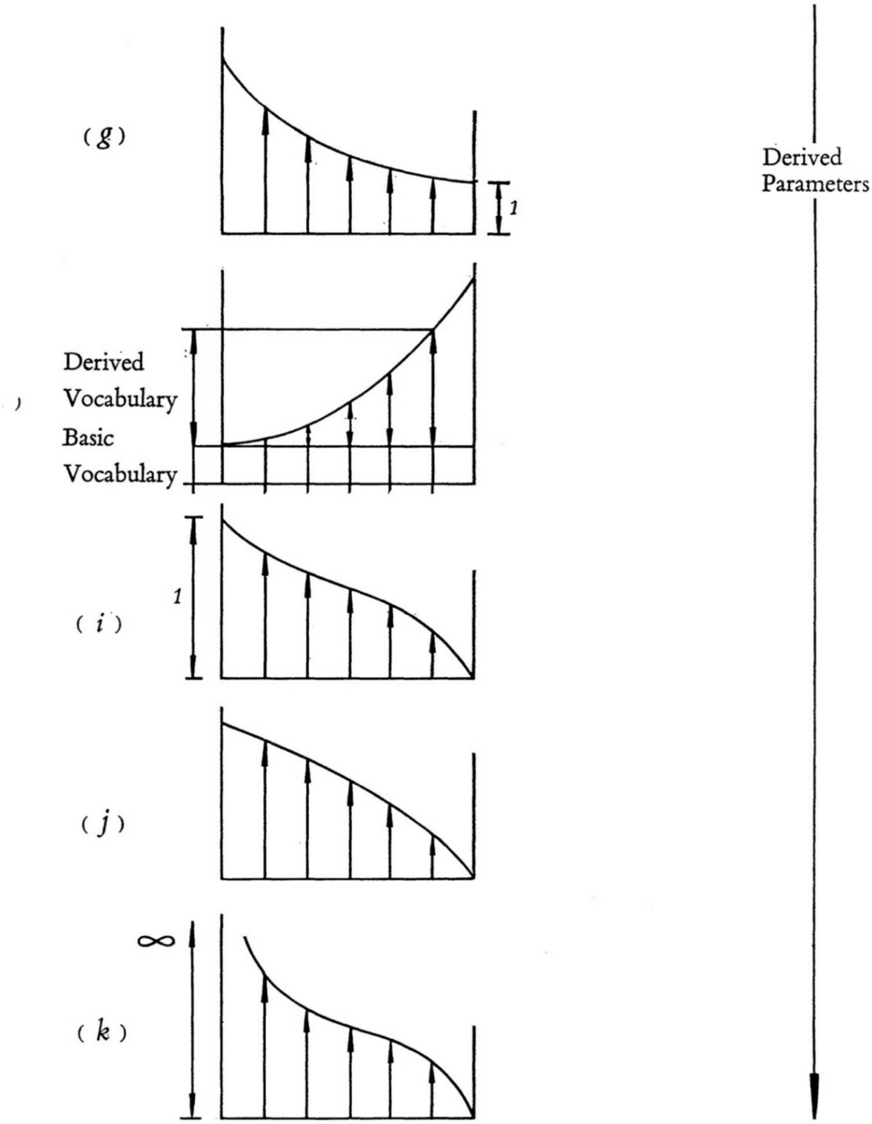


FIGURE 4. *(d)* Potential depth of indexing and retrieval; allowable diffuseness of information; number of terms per item of information. *(e)* Permutability of retrieval criteria. *(f)* Degree of hierarchial definition of indexing.

single indexing "term," since it designates only one particular place in the classification.

It will be noticed that this classification is an (unpermuted) multiple-criterion classification, having *structure* as its main criterion, *fuels* as sub-criterion, and *operational troubles* as a sub-sub criterion. In practice, most classifications will not have more than two indexing criteria, so that the possibility of indexing at will by additional criteria such as *operational troubles* is lost at this end of the continuum.

Derived mathematical parameters

In addition to the main or defining parameter, average term length, the descriptive continuum possesses a number of "derived" parameters which are functions of the main parameter.

The most important derived parameter is the *potential depth of indexing*. This parameter, which is proportional to the number of potential index criteria per item that can be used, is shown schematically in Fig. 4*d*. For a pure classification system this value is unity, since one cannot index by a larger number of indexing criteria than those that have specifically been built into the classification. As we move away in the continuum and shed the constraints of classification systems discussed in earlier sections, larger numbers of indexing criteria can be utilized without undue penalties in cost and complication. When we reach the other extreme in the continuum almost unlimited indexing depth becomes practical.

This and other parameters show that, in general, the more diffuse the information, the more advisable it is to keep to the short-term end of the continuum.

Indexing depth, however, involves more than just the number of indexing terms that can be combined for a search. A related derived parameter is *permutability of retrieval criteria,* schematically shown in Fig. 4*e*. In a classification for retrieval by one indexing criterion only, this must be the criterion on which the main classes are based; for retrieval by two criteria these must be the criteria on which the main classes and first sub-classes respectively have been based, etc. We cannot retrieve by a single criterion which forms only the sub-classes, or the sub-sub-classes. At this end of the spectrum, the permutability is therefore equal to zero. On moving toward the other end of the spectrum more and more permutability of retrieval criteria is obtained, until at the short-term end of the continuum, full permutability is available.

Together the two parameters discussed above determine the accuracy with which information can be indexed and the detail in which information can be retrieved. We will therefore lump together these parameters in the term *retrieval power*. This is the most important single feature of an indexing system.

A third derived parameter is what can best be termed *hierarchial definition of indexing*. (See Fig. 4*f*.) In a classification system this parameter is equal to one. That is to say, there is a 100 percent hierarchial definition, since to index by any one criterion is to index by all possible hierarchial levels. In our previous example of an item of information on the "erosion of turbine blades in bunker-C grade oil burning jet engines," each of the three possible indexing criteria "structure," or "fuel," or "operational troubles" was indexed by all possible hierarchial levels: the structure as "turbine blades, turbines, steady-state engines, internal combustion engines"; the fuel as "bunker-C grade oil, mineral fuel, liquid fuel." Operational troubles could have been indexed as "erosion, wear," if the classification had provided for the possibility of a third indexing criterion.

At the other end of the continuum, the area of key word indexing, hierarchial definition disappears out of the "retrieval apparatus." By using our previous example again and indexing this document by "structure," it then becomes more or less arbitrary whether to index simply by the term "turbine blades" or by the next higher level such as "gas turbines" or by a still higher level or any combination of these.

In selecting index terms there are only two degrees of freedom, namely the *index criterion* we use and the *hierarchial level* within this criterion. The first degree of freedom corresponds to the combination *retrieval power* shown in Figs. 4*d* and 4*e*. The second corresponds to the *degree of hierarchial definition* shown in Fig. 4*f*.

Although no exact mathematical relationship is as yet available, it seems clear that *greater hierarchial definition can only be obtained at the expense of retrieval power, and vice versa*. This must be considered the first basic law of the descriptive continuum. If X designates the "retrieval power" and Y designates the hierarchial definition, it could be roughly expressed as follows: $X \cdot Y =$Constant.

When a search has to be performed in a system located at the short term end of the continuum where the hierarchial definition is low, a complete search may involve a number of individual searches using terms of different hierarchial level. In a search by two terms of different indexing criteria, each having a number of possible hierarchial levels, it is necessary to search by each of all possible permutations. In contrast with this, at the other extreme of the continuum, the area of pure classification, only one search is required. The number of individual searches required for a complete systems search is schematically shown as a parameter in Fig. 4*g*.

Another parameter of importance to the economics of the system is the size of the *access apparatus*. This is indicated in Fig. 4*h*. At the short-term end of the continuum the access apparatus consists merely in the alphabetic list of the smallest single words used for indexing. This basic vocabulary is usually very small.

At the long-term end of the continuum, the access apparatus includes combinations of these single words. The number of these combinations increases greatly as the combinations get longer and longer.

Fig. 4*i* shows the need for a *mechanical coordination* or *coincidence mechanism* as a function of the main parameter, average term length. At the short-term end of the continuum, a coordinating mechanism is indispensable. At the other extreme end no such mechanism is required. In the middle of the continuum there may be a need to search for the coincidence of subject headings. Without a coordinating mechanism, the potential depth of indexing and retrieval shown at the short-term end in Fig. 4*d* remains a *potential* only. The same conclusion holds for all other parameters. The actual values of the parameters which are obtained will otherwise be far less favorable.

Fig. 4*j* shows the *possibility of false coordinations* of two or more terms as a function of the average term length. As explained in the previous section, it is zero at the long-term end of the continuum and a maximum at the short-term end. False coordinations do not result in loss of information, but in the retrieval of some non-pertinent documents among all the required pertinent documents. It is a "noise" factor usually having a minor nuisance value only. This parameter is therefore of little practical importance.

Figure 4*k* shows the *capacity for handling semantic indeterminacy,* that is, the capacity of an indexing system for indexing in areas where a sharply defined and well-standardized vocabulary is not yet available. As explained previously, classifications cannot handle this type of "frontier" information at all. The closer the position to the short-term end of the spectrum, the greater the capacity to handle semantic ambiguities or indeterminacy. At the extreme short-term end, it increases to an almost infinite degree, and even the vaguest concepts can be described by putting a number of different terms together to form a kind of telegram-style description of the concept. Any one or any combination of these terms can then be used to retrieve the information item.

The general theory of indexing as presented above is open to mathematical treatment. The parameters presented so far are mathematical in nature. The curves shown are of course only schematic indications of the true mathematical functions. Determination of their exact shapes will have to await the establishment of the correct mathematical relationships. A complete mathematical interpretation of this general theory is planned.

Cost parameters:

In addition to these operational parameters a number of others could be listed. These would be *cost* parameters. Although the full mathematical development has not yet been achieved, certain general conclusions seem possible.

1. *Cost of indexing*. It is lowest on the short-term end of the spectrum and highest on the long-term end; the closer to this end the indexer operates, the greater will be his need to consult the existing access apparatus, and the higher will be the indexing cost.
2. *Cost of access apparatus*. This parameter will follow roughly the same pattern as the indexing cost.
3. *Search cost*. A pure hierarchic classification should in theory allow immediate retrieval by a single search, with no required mechanical searching aid. This makes for low search cost. On the short-term end of the continuum every search question may involve a number of individual searches, and some form of search machine is required. Higher search costs can therefore be expected. However, this factor can be reduced in significance by the choice of a proper coordinating mechanism, as will be explained in a subsequent paper.

As far as the total costs are concerned, the general conclusion seems to indicate that normally the overall cost of operating a system, that is the total *input* cost (cost of *entering the data*) and total *output* cost (cost of *searching*) will, in most cases, be considerably lower at the short-term end.

SELECTION OF POSITION IN THE CONTINUUM

When we have information of *low diffuseness,* which does not require more than one or two different indexing criteria, and when these criteria have always the same hierarchial subordination, we can in theory operate at *any* point in the continuum and use either *classification* or *keyword indexing*. Generally speaking, however, the cost will be much lower at the short-term end, so that we will normally select this end of the continuum. One possible exception is the case in which the input load is very small and the search load very high. In this case the low indexing cost of the short-term end is of little help and the low search cost of the long-term end would definitely be desirable.

For information of *high diffuseness* for which *permutability* of the indexing criteria is required, we have no choice but to operate at the short-term end of the continuum regardless of the ratio of input to output load, as it is only at the short-term end that full permutability and the highest possible indexing depth are available.

The present paper represents the first attempt to commit this general theory to print. We are fully aware that this first writing may still be far from complete and may contain many weak points and inconsistencies in its presentation. But we hope it will contribute to a better understanding of the problems of indexing and may serve as a foundation upon which in time a fully mathematical theory of indexing can be erected.

Algebraic Representation of Storage and Retrieval Languages

R.A.FAIRTHORNE

The fundamental elements of any system for storage and retrieval are the entries, bookings, or postings which ultimately connect specifications of requirements with the stored items. At any moment the system may be regarded as a double-entry table with components of specification naming the rows, say, and component of stored items naming the columns. Existence of a posting at the intersection of a row and a column denotes that the corresponding component of specification should invoke the corresponding component of stored item. Thus, given the terms in which requests must be specified and those in which items must be marked and the existence and non-existence of postings, this table describes completely the classificatory or taxonomic structure and properties of the system as it stands at the moment. That is, it shows what you should get when you ask for what; not the way in which you get it.

If, additionally, the double-entry table gives numbers representing such things as frequencies of requests using various terms, it will then describe completely the statistical "informational" behaviour of the system, regarded from outside as a data-processing "black box." From this point of view any further information is irrelevant. Either it is about the particular means by which the retrieval system achieves the current input-output behaviour or it is about the rules by which new postings are assigned when new items enter the system. The former is beyond the scope of black-box theory; the latter is at the moment beyond its powers.

Even when black-box theory becomes strong enough to cope with growing systems, that is, with intensional or "semantic" information as well as with extensional or "selective" information, it will concern itself only with the assignment and distribution of postings. It cannot concern itself with various ways in which they can be written down, partitioned, and manipulated.

In this sense, the retrieval system is a set of rules for assigning postings that connect textual content with descriptive terms, both terms and content being

R.A.FAIRTHORNE Royal Aircraft Establishment, Farnborough, England.

assumed determined or determinate. The system can be realized in practice in as many ways as ingenuity suggests or means allow. However, if one is concerned both with what one gets when one asks for what, and how to get it, the way in which postings are written and manipulated is of the same stature as the way in which they are assigned. A system without means of retrieval is as useless as means of retrieval without a system. Retrieval devices can act only on marked material objects in ways depending on the marks, the material objects being partitions of the double-entry table of postings, the marks being the postings written in some form acceptable to the device. If, for instance, the available clerical devices are suited to finding the common membership of given lists, we partition the double-entry table along the rows (e.g., as “aspect” or “feature” cards) and write the postings in a way congenial to the comparing and recording organs (e.g., as an aperture or a microprint in a “dedicated position”). If the system demands both common membership of given lists and lists of total membership, partitioning along both rows and columns should be possible, and the postings must be written in some at least ternary (three letter alphabet) script. At all times the way in which postings are written, or into which they may have to be translated, must reconcile the clerical means of retrieval and the fundamental semantic structure of the system. Subject classification by itself can retrieve nothing.

Characteristics of a working system therefore reflect the way in which the postings are spelt and manipulated as well as the structure of what they refer to and how they refer to it. For example, the marked physical differences between predominantly “analogue” and predominantly “digital” calculating devices are due entirely to the ways in which they represent numbers. Also the difficulty of classifying some devices either as “digital” or as “analogue” reflects the fact that there are as many methods of naming numbers, intermediate between “analogue” (repetitive) and “digital” (combinatorial), as one cares to think up.

A clear-cut example of the way in which one talks about things dominating the characteristics of the things talked about is the “law of logarithmic distribution of initial digits.” Examine almost any extensive collection of numerical data, such as tables of statistics, that are written in variable-word-length decimal radix form; that is, with the usual convention that 0 is never used initially and therefore, of necessity, using an eleventh character, the “space,” to mark the end (or beginning) of a number word. About a third of the entries will be found to begin with a 1; about a sixth to begin with a 2. In general, the fraction of entries that begin with a d is approximately $\log_{10} [(d+1)/d]$. This has been, and sometimes still is, interpreted as a law of nature governing the distribution of magnitudes in the physical world. If so, the law must be awkward to enforce, because the distribution changes if the observer chooses to write down the magnitudes in duodecimal, say, instead of decimal radix notation. If he chooses to use binary radix variable-word-length notation every word begins with a 1. In general, if you use radix *r,* the fraction of entries that begin with a *d* is approximately $\log_r [d+1)/d]$.

This is not just a manifestation of the known fact that a straight-line law connecting any empirical data always can be achieved with the aid of suitably scaled logarithmic paper and a robust conscience. A little consideration shows that in a run of consecutive integers from one to some power of ten, initial digits are distributed uniformly over 1, 2,· · ·, 9. If the run does not stop at a power often, the lower initial digits must predominate. Deeper analysis shows that with reasonably large samples, repetitions being allowed, from a large enough initial segment of the integers, the logarithmic relation quoted above is a good approximation (3, 4). The approximation is remarkably insensitive to bias in the sample, the sensitivity vanishing as the radix of representation decreases to two.

In short, this logarithmic law is a direct consequence of our spelling rules for number names. Not obedience, but disobedience, is evidence for particular extensional characteristics of the set of numbers named or the method of selecting them. Similar relations would be expected, and are found, with any vocabularies built up combinatorially from alphabets, phrases built up from vocabularies, or any structurally equivalent naming system; e.g., bibliographies built up from references. The best known and studied is the Zipf-Mandelbrot relation, e.g., Brillouin (2), between word-rank and word-frequency in ethnic languages. Since phrases in ethnic languages are effectively number words of low radix, this relation is significantly affected only by such pathological characteristics of language as restricted vocabulary and elaborate syntax (Hebrew, Basic English, children's talk) and unrestrained word-order (as in some mental disorders and many of the mathematical models used to derive the law theoretically). The proportions of very common and very rare words do, however, cause local but significant deviations from the general relation. (The latter needs only the hypotheses that, on the average, less costly signals are used oftener than more costly; and that signals of equal cost are used equally often.)

Important to retrieval are the Bradford law of scattering of bibliographic references and its implications (1, 5). It deserves attention and extension, for it may apply to greater or lesser extent to, for instance, the distribution of postings over “aspect” or “feature” cards and, conversely, the incidence of “aspect,” “features,” or to some extent even the more fundamental types of descriptors attributed to items. If the last are used as an index of subject matter, possibly only their deviations from some Bradford-Zipf-Mandelbrot relation are significant.

Many other clerical activities are affected as much by spelling rules of the names as by properties of the nominata. For instance, sorting and, a fortiori, marshalling. Sorting is essentially segregation of all items that are homonymous with respect to certain varieties of names. Unless the radix or size of alphabet of the names is not less than the number of varieties into which the items are to be segregated, sorting must be iterated according to partial marks, usually the characters of successive digits. Clearly the efficiency of the process, in terms of number of passes or of loading of pockets, is as vulnerable to the vocabulary of marks as it is to the composition of the collection to be sorted. The latter is outside one's control and, usually, one's knowledge. The former is not.

Marshalling, which entails assembly as well as segregation, is even more at the mercy of the vocabularies of marks. One has only to compare the (hypothetical) task of marshalling decimally marked wagons on a trident with the same task with the wagons marked in ternary radix notation. The marks, operations, and facilities then have a common structure. Marshalling may be regarded as a particular case of transcription from a vocabulary in which the words represent initial rank to another in which the same words represent the final rank.

Some of the quoted examples deal with combinations of symbols and operations, some with permutations, some with both. Structurally there is no fundamental difference; permutations and combinations are reciprocally related according to interpretation. For instance, the apertures in a Taylorstencil “aspect” card can be usefully interpreted as either a particular combination of items, or as the binary radix representation of the serial number of that combination in a certain enumeration, unpunched sites being interpreted as 0's and punched sites as 1's. The enumeration is made by associating unity with citation of the first item, two with citation of the second, four with citation of the third and, in general, 2^{n-1} with citation of the *n*th item. The serial number of a particular combination of items is found by adding the numbers associated with the components. The entire enumeration is obtained from the expression

$$(1+a_1 x^{2^0})(1+a_2 x^{2^1})(1+a_3 x^{2^2})\cdots(1+a_n x^{2^{n-1}}) \qquad (1)$$

where a_r denotes citation of the *r*th item.

On multiplication we get terms of the type $a_p a_q a_r \cdots x^{2^{p-1}+2^{q-1}+2^{r-1}+\cdots}$ in which the coefficient represents the combination of items and the exponent its serial number. Proof that each of the 2^n combinations has a unique serial number is elementary.

But expression (1) is not essentially different if we write it as

$$(0+1\,x)\,(0+1\,x^{2^1})\,(0+1\,x^{2^2})\cdots(0+1\,x^{2^{n-1}}) \qquad (2)$$

order being preserved in multiplication so that, for instance, 1 in the third term denotes occurrence of a 1 in the third binary place. On multiplication we get terms such as 00· · ·01101 x^{13}, the coefficients being the name of the exponent in binary radix constant-word-length notation.

Suppose the aspect cards were only just large enough to accommodate the particular combinations written on them; e.g., lengths of perforated tape trimmed after the last entry. They then could be interpreted as the serial numbers of the combinations expressed in variable-word-length binary radix notation. Either interpretation has the structure represented by

$$\cfrac{Bax}{1-\cfrac{x(b+ax)}{1+x(b+ax)-\cfrac{x^2\,(b+ax^2)}{1+x^2\,(b+ax^2)-\cdots\cfrac{-x^{2^n}\,(b+ax^{2^n})}{1+x^{2^n}\,(b+ax^{2^n})-\cdots}}}} \qquad (3)$$

B denotes either the beginning of a new combination or a space, that is, a special character or characters indicating the beginning (or end) of a new word. Note that there is now no coefficient of x^0 (i.e., of unity) in expression (3) for we can have neither an aspect card for an empty aspect on the one hand, or name for zero on the other. To deal with these the system would have to provide unsystematic treatment, represented by the addition to (3) of a constant term not deducible from the structure of the rest of the expression.

This is an unlikely form for a Taylor system, but as soon as such a system demands more than one aspect card for an aspect its structure has to reflect characteristics of both (1), or (2), and (3). The words now vary in length, but not by one letter at a time. Elaboration of this would merely complicate the present discussion.

Suppose now that the cards are able to display more than two states at a site. For instance, the references might not be to individual papers but to the number of papers by individual authors, or to whether they had written for or against a particular aspect, or not at all. In the last case, the card must be able to display three distinct states at each site; in general, it will be required to assume any one of r_1 distinct states in the first site, r_2 in the second, and so on. The structure is then represented by

$$(1 + a_1 x + a_1^2 x^2 + \cdots + a_1^{r_1 - 1} x^{r_1 - 1})(1 + a_2 x^{r_1} + a_2^2 x^{2r_1} + \cdots + a^{r_2 - 1} x^{(r_2 - 1) r_1})$$
$$(1 + a_3 x^{r_2 r_1} + a_3^2 x^{2 r_2 r_1} + \cdots) \cdots (1 + a_s x^{r_{s-1} r_{s-2} \cdots r_2 r_1} + a_s^2 x^{2 r_{s-1} r_{s-2} \cdots r_2 r_1}$$
$$+ \cdots + a_s(^{r_s - 1} x^{(r_s - 1) r_{s-1} r_{s-2} \cdots r_{s-1}}) \tag{4}$$

or

$$\frac{1 - a_1^{r_1} x^{r_1}}{1 - a_1 x} \cdot \frac{1 - a_2^{r_2} x^{r_2 r_1}}{1 - a_2 x^{r_1}} \cdots \frac{1 - a_s^{r_s} x^{r_s r_{s-1} \cdots r_2 r_1}}{1 - a_s x^{r_{s-1} \cdots r_2 r_1}} \tag{5}$$

Both (4) and (5) are the same initial segment of the expansion in positive powers of *x* of

$$1 / (1 - a_1 x)(1 - a_2 x^{r_1})(1 - a_3 x^{r_2 r_1}) \cdots (1 - a_s x^{r_{s-1} \cdots r_2 r_1}) \tag{6}$$

If we concern ourselves with $a_1, a_2, \ldots a_s$, only (that is, if we put all a_t, $t>s$, equal to zero), the relevant segment will be the first $r_s r_{s-1} r_{s-2} \cdots r_2 r_1$ terms of the expansion, which will give all the selections as coefficients and their unique serial numbers as exponents of *x*. This is also the greatest number of distinct states the aspect card can display, so the coefficients of higher powers represent physically unrealizable states, because they involve factors of form a_p^t, where *t* is greater than $r_p - 1$.

Expression (6) generates upon expansion the representation and enumeration of any set of selections allowing up to r_1 things of the first kind, r_2 of the second, and so on.

This may be interpreted, not as an extension, but as a degeneration of the one-or-none binary selection represented by (1). We may lack either the means or the need to distinguish between separate items of the same more general type; for instance, between individual papers by the same author or, when we are concerned only with the number of letters in a word, between the different sorts of letters (other than spaces or their equivalents).

Suppose that, in expression (1), a_1, a_2, a_3, say, represent individual papers by the same author. Only four states now need to be enumerated, (no-paper through three-papers), instead of eight (the number of ways of making selections from three distinct objects). The only powers of *x* needed for the enumeration are, therefore, zero through three. So, in the first instance, we can write the appropriate terms of (1) as

$$(1 + a_1 x)(1 + a_2 x)(1 + a_3 x) \tag{7}$$

On multiplication the coefficient of x^2, say, is

$$(a_1 a_2 + a_1 a_3 + a_2 a_3) \tag{8}$$

But we are not able to distinguish between a_1, a_2, a_3, so this must become

$$(a^2 + a^2 + a^2) \tag{9}$$

where a^2 represents the statement "There are two papers written by author *a*," because we can no longer say which the two are. Making this statement three times over is the same (in the sentential calculus) as making it only once. So, because of our reduced discrimination, (9) finally becomes

$$(a^2 \vee a^2 \vee a^2)=a^2 \qquad (10)$$

where v is to be read as the symbol for disjunction of statements.

When the *a's* have suitable interpretations, such as patterns that can overlap or coincide, we would use ∪, the symbol of set addition, instead of *v*. Whatever the interpretation, or the symbol, it must formally reflect the type of blurring due to our coarsening the discrimination. The plus sign, +, used in these expressions denotes simply that the items separated by it are distinguishable somewhere within the system by means available to the system, but not necessarily that they are distinguished. For instance, we might be interested only in the number of papers by given authors, but be compelled to use Taylor cards that distinguished between individual papers. Then we would, as before, equalize appropriate exponents (serial numbers of component items) of *x* in the factorized expression (1) because their coefficients are not to be distinguished in the new enumeration. Nevertheless they remain distinct patterns on the card, and will be so read by the usual sensing devices. So we cannot simplify (8) to (9). Such bracketed terms are the synonyms of the corresponding serial numbers (exponents of *x*), and the sensing device must have some means for dealing with such synonyms, a complication reflected by the structure of (7) being more complicated than the structure of the expression using relations such as (10); to wit

$$1+ax+a^2x^2+a^3x^3 \equiv \frac{1-a^4x^4}{1-ax} \qquad (11)$$

If the sensing means detected a single, a pair, or a triplet of *a's* without taking into account the identities of the individual items, (9) would have to be written

$$3a^2 \qquad (12)$$

because the pair response would have to be given to three distinct patterns. The expression representing this part of the system is then

$$(1+ax)^3 \qquad (13)$$

This is the least complicated expression possible if the original card, or system, distinguishes between each item and you must extract only the number of items of given sorts. That is, if the system contains synonyms with respect to the characteristics that interest you.

The coarsening of discrimination detailed above is compactly displayed by writing the relevant factors of (1) in the form

$$\frac{1-a_1^2x^{2^1}}{1-a_1x^{2^0}}\cdot\frac{1-a_2^2x^{2^2}}{1-a_2x^{2^1}}\cdot\frac{1-a_3^2x^{2^3}}{1-a_3x^{2^2}} \qquad (14)$$

This becomes (7), which can be written

$$\frac{1-a_1^2x^2}{1-a_1x}\cdot\frac{1-a_2^2x^2}{1-a_2x}\cdot\frac{1-a_3^2x^2}{1-a_3x} \qquad (15)$$

The next stage, (13), can be written

$$\left(\frac{1-a_1^2x^2}{1-a_1x}\right)^3 \qquad (16)$$

The final stage is

$$\frac{1-a^4x^4}{1-ax} \qquad (11)\text{ bis}$$

In general, confounding groups of items produces the factors of expression (5). Therefore the parent function is, as asserted above, a degenerate case of the denominator of (1) when written in the form of (14). From the latter can be produced, by coarsening of discrimination and truncation of the power expansion (i.e. reducing the range of the vocabulary), all radix type representations on the one hand or selections with repetition on the other.

The denominator is thus the basic generator of an indefinitely large family of finite vocabularies having some structural characteristic in common. We should note that the first *n* terms of such a vocabulary are given by an expression of the form

$$\sum_s \frac{A_s(1-a_s^n x^n)}{1-a_s x} \qquad (17)$$

where the summation is over the roots, $1/a_s$, of the denominator considered as a polynomial in *x*. The *A's* are functions of the roots only, and are found by the usual Euclidean division algorithm.

This partial fraction representation is the basic representation of vocabularies and will be arrived at eventually however tortuous the route; e.g., by detour through finite difference equations representing the recurrence relations between coefficients of successive powers of *x*.

The word corresponding to the *n*th item of the vocabulary is thus the sum of the coefficients of x^n in the expansions of the partial fractions. For arithmetical features of words, such as relative frequency or number of synonyms, or when the word can be regarded as coexistence of distinguishable parts, the summation can be represented by the plus sign. Other occasions may demand other forms of addition; ordinal, cardinal, set, lexicographic, etc.

The discussion has been illustrated in terms of Taylor cards. It is equally illustrated by the number representations on a pinwheel desk calculator, of which Taylor cards are irrevocable versions, structurally. Here a displayed configuration, 0282 say, is to an observer who can recognize the characters one of the permutations, with repetitions, of the ten numerals 0 through 9. nternally, it is a particular selection which has combined no teeth on the thousands' wheel with two teeth on the hundreds' wheel, eight on the tens', and two on the units'. To a certain degree the two vocabularies are structurally the same. We can substitute a_3^8, say, for 8 in the coefficient of x^{80} and, in general, we can equate coefficients of like powers of *x.* We can do so only because the mechanism is a physical operator for physically equating coefficients. The vocabularies are structurally the same, but the alphabets are not. Users would resent having to count the number of teeth rather than to recognize the numerals; they would also resent having to pay for a character reader to recognize the numerals rather than for pinions to count the number of teeth. Words are composite (many-lettered) or irreducible (letters of an alphabet) relative to a particular means of observation; the properties are not absolute. Also they are relative to the spelling rules. So long as n+1 is represented by means of one more tooth, and one less blank, than the representation of *n,* the tooth-setting device is a ten-word vocabulary using a dyadic (two-letter) alphabet. If the assignment of teeth were arbitrary, so that the number of teeth representing n+1 could not be deduced from that for *n,* it becomes a ten-word vocabulary formed from single letters of a ten-letter alphabet. The mechanism for equating coefficients is correspondingly complex. More elderly computors may recall the weird cam-operated keyboard dials of the first motorized Mercedes Euklid calculaters.

The *x's* themselves may be the names of some other nominata, instead of being purely the vehicles of some enumeration. Often they represent the input unit pulses. The mechanism described cannot distinguish between n and n+10000k, where k is any integer. Therefore the factor $(1-x^{10000})$ must be inserted in the denominator in order to provide the appropriate homonyms. A negative power in the numerator will deal with whatever convention is used for negative numbers.

So far we have shown: first, that the properties of the way in which we talk about things can distort or obscure the properties of the things talked about. Second, by possibly tedious discussion of simple examples, that the various ways in which clerical systems use signals within a system are greatly simplified if the various notations and operations have structures easily derived from some common abstract structure.

Therefore, to separate, analyze, or synthesize clerical aspects of retrieval we need some model which will represent vocabularies (i.e., ordered or partially ordered sets of words) in structure as well as in extension, to any degree of approximation, discrimination, or range. Such models should cope with extensions of vocabulary by creating longer words through multiplication or repetition (exponentation of alphabets and vocabularies), the multiplications being cardinal, ordinal, or, in general, lexicographic. Also they should cope with extensions of vocabulary by phrase-formation; that is, using the vocabulary as the alphabet of higher vocabularies by means of hierarchies of spaces, which are characters which, by rule or the nature of their representations, cannot occur adjacently (e.g., a duration greater than some given minimum and, perhaps, less than some maximum which is less than twice the minimum). The models should deal quantitatively with magnitudes that are additive over word or phrase, such as length in terms of number of digits or of physical length or duration. Also with magnitudes multiplicative over word or phrase, such as relative frequency when letters or words are independent. For retrieval notations partial or simple orderings and set operations on characters must be coped with. Finally, the overall algebraic manipulations of the representation should have clerical interpretations in terms of discrimination, matching, segregation, transfer, or retention.

This is a large order, and no model is likely to satisfy all of it all the time. It will be quite useful if it satisfies some of it some of the time. Such a model is best regarded as a process of generating either samples or the whole structure of the vocabulary with such detail as is necessary or possible. Communication theory favours the stochastic model, which generates samples, because the statistical features of its material are usually the only ones accessible to it. Retrieval systems deliberately use (or create) systematic features in their material and devise systematic notations and operations to reflect them. For this the best model is an expression that can generate the systematic vocabulary and in itself display the general structure. The more systematic the orthography and assignment of magnitudes, the closer an approximation such a model will be.

In the course of this discussion quite a promising type of model has emerged, taking into account only the orthography, not the numerical magnitudes that may be associated with the words. We will now look at these. Already we have seen that it is consistent to take, for instance, the 3 in (12) as indicating that the a^2 can occur in three ways. In general, such coefficients indicate the number of times the word has occurred or the relative frequency of its occurrence in a particular population of words from this vocabulary. If the coefficient is the relative frequency, then the relative frequency of joint occurrence, or product, of two independent words is the product of their several frequencies.

If we list only the relative frequencies of the words of a vocabulary, without representing their spelling, we can write it as

$$V = p_1 x^1 + p_2 x^2 + p_3 x^3 + \cdots + p_n x^n + \cdots$$

(18)

p_n being the relative frequency of the nth word.

The series can be represented by a continued fraction of form

$$V = \cfrac{a_0 x}{1 + b_1 x - \cfrac{a_1 x^2}{1 + b_2 x - \cfrac{a_2 x^2}{1 + b_3 x - \cdots}}} \qquad (19)$$

This is, with changed notation, the standard *J*-fraction expansion of the series. There are many algorithms for computing it, e.g., Wall (6). All basically are the Euclidean G.C.D. algorithm applied to polynomials, corresponding exactly to the algorithm for purely numerical continued fractions.

When (19) is evaluated by known means as far as the nth partial denominator, $1 + b_n x$, we get a fraction whose numerator and denominator are polynomials of degree *n* in *x*. Expansion of this fraction by long division yields a power series that coincides with (18) through $p_{2n} x^{2n}$. We can therefore obtain a polynomial fraction that approximates the frequencies to as many terms as we like.

For this procedure to work, the *p's* must satisfy certain broad conditions, besides being all non-negative. In general you cannot use it to obtain some special structure by systematically putting some *p's* equal to zero.

Next, the fraction can be represented by *n* partial fractions as in (17), the roots, $1/a$, of the denominator being now numbers. The entities taken as roots in (17) are letters in some generalized sense. Study is needed to show how and when they become clerically realizable.

Each partial fraction yields, on expansion by long division, a vocabulary whose *r*th word has relative frequency a^r. *V* is the sum of such vocabularies so p_r is the sum of the *r*th powers of the *n* roots for all *r* not greater than 2*n*. This is therefore the solution of the moment problem for any finite segment of *V*.

Formally, the same applies to the alphabetical roots of (17). If any of the *a's* can be interpreted as an available character (sign-design), the denominator containing it generates the simple repetitive vocabulary whose *n*th word is *'aaaa· · ·*(*n* times) preceded, or followed, by a space character, which is here an unmarked area of paper of indefinite extent. Nevertheless, it is a character, because a definite choice is needed to produce it as an alternative to an *a*. All signalling systems must have alternative choices, from their nature, and an *r* character alphabet offers at most $r-1$ choices. Thus all alphabets must have at least two characters, one of which must be of space type (unless it is known that the alphabet is to be used once only for one one-word message).

Such repetitive synonym-free dyadic vocabularies are clearly the simplest of all types of vocabulary. They are, of course used explicitly by discrete (digital) analogue systems, and implicitly by non-discrete (continuous in the sense of blurred) analogue systems.

Thus the individual words and associated quantities of an ordered vocabulary are formally equivalent to the moments, numerical and alphabetical, of some function of order, of which the partial fraction representation, as in (17), is the discrete Stieltjes' transform. We will therefore refer to expressions of this type as "Stieltjes' representations" or "Stieltjes' transforms" of the vocabularies. They are closely allied to the discrete spectra, autocorrelation functions, and other representations used to analyze the physical aspects of signals.

The value of Stieltjes' transforms of vocabularies is no less because the dyadic vocabularies are not directly realizable; e.g., a character may be negative, formally or numerically. The original denominator may be reducible to alphabetical roots that are real for some clerical devices, but not for others. How many, what kind, and by what means they are to be realized should be a profitable field of investigation.

Though these and other fractions have been said to generate vocabularies by long division, it must be recalled that this phrase is purely the mathematical interpretation of a clerical routine for making marks in a certain way (writing a string of *n a's* as a^n is in this context merely an inessential abbreviation). Multiplication means only that one makes a list of all combinations of marks, one from each factor, and orders this list according to the rules for the particular type of multiplication. Differentiation is meaningless here in terms of an approach to a limit; in combinatorial analysis, as here, it is a discriminating operator as its name implies, e.g., $a\,(dx^3/dx)$ lists all ways of substituting an *a* for an *x* in *xxx;* to wit, $axx+xax+xxa$, which may sometimes be abbreviated legitimately to $3ax^2$. Similarly, the derivative of a polynomial cannot here be interpreted as the slope of a curve, but is the remainder after two successive divisions by a suitable divisor (i.e., the confluent divided difference). In short, mathematical algorithms, including computer programs, are clerical algorithms for making marks. Mathematical concepts are concepts, and therefore not accessible to clerical operations though applicable to them on occasion. Data processing systems, for instance, can handle neither numbers nor information, they can handle only signals, which are physical events. Numerical and informational concepts, however, can or should be handled by those who handle data processing systems. We are therefore entitled to use, for instance, clerically unrealizable repetitive vocabularies so long as they originate in and go to form consistent and useful clerical results.

So far the exponents of *x* in such expressions as (18) have denoted rank in general. Words often may have to be ranked by some property of words, such as the number of letters or, in general, the cost. This can be represented by an extra coefficient u^C, say, whose exponent is an integer representing the cost. The costs are assumed to be positive integral multiples of some common unit. With this representation, if the cost of a word is the sum of the costs of the letters of which it is the product, the exponent of u in the word will be the cost of the word. Other additive costs can, when necessary, be represented by the exponents of other extra coefficients.

For instance, the coefficient of $u^C v^n$ in the expansion of

$$V = \frac{B\left(p_1' a_1' u^{c_1'} + p_2' a_2' u^{c_2'} + \cdots p_{r-1}' a'_{r-1} u^{c'_{r-1}}\right) v}{1 - \left(p_1 a_1 u^{c_1} + p_2 a_2 u^{c_2} + \cdots + p_{r-1} a_{r-1} u^{c_{r-1}}\right) v} \quad (20)$$

is the list of all N-letter words costing exactly C, together with their relative frequencies, that can be formed from an alphabet of r letters (including the space B and special initial characters). The number of such words is found by putting the p's and a's equal to unity. If we are not interested in the number of letters and the costs simultaneously, we can put v equal to unity. Then all N-letter words will be given as the coefficient of u^n after the letter costs, the c's, have all been made equal to unity. The number of such words will be found by putting the a's equal to unity also.

It is easy to see that the denominator of (20) will the have only one positive root, lying between 1/2 (for an infinite alphabet) and 1 (for a dyadic—one letter and one space—alphabet). The repetitive vocabulary generated by the corresponding partial fraction models the average relation between rank and cost; the vocabularies generated by the other roots correspond to oscillations about the average. This is true whatever the numerator, and therefore whatever the length of segment of V considered. The amplitudes and similar parameters do vary with the range of vocabulary considered, which manifests itself in the numerators of the partial fractions. In (20) the primes in the numerator denote frequencies and costs of initial letters or letters used initially. This gives a slightly better model, but does not affect the structure or roots of the basic representation.

In turn, we can use V as an alphabet from which phrases are generated, just as the alphabet of V is generated as a vocabulary of words formed of sub-alphabetic components, if we have a suitable facsimile device that accepts these components as an alphabet. If all such phrases are permissible, frequencies multiplicative, and costs additive, the phrase-generating expression is

$$V_2 = \frac{B_2 V'}{1 - P_1 B u^{k_1} V} \quad (21)$$

where B_2 is a second order space (e.g., a stop) of indefinite cost; $P1$, k_1, the frequency and cost of the first order space B; V' the vocabulary, with frequencies and costs, of initial words or words used initially.

This hierarchy can be extended upwards or downwards indefinitely. Always, however, it can be represented as the sum of dyadic repetitive alphabets (of the same hierarchic order).

This paper has outlined a possibly useful method of representation in which vocabularies and hierarchies of vocabularies are regarded as the sum, in any consistent sense, of repetitive dyadic vocabularies, not necessarily clerically realizable. It generalizes and unifies many special methods, such as various algebraic identities used traditionally to demonstrate number representations, and models used in investigating some properties of ordinary language. Here it can be used to discover the sympathetic magic principles (like-produces-like) characteristic of linguistic systems and to apply them usefully to more systematic vocabularies. We have seen that to some extent it can cope, though not simultaneously, with additive and, in general, modular properties such as cost and selective information, and with the partial orderings and looser generalized operations, synonymity and homonymity, that are essential to retrieval.

The next stage is to deal with the vocabularies of operations for transforming vocabularies. These are the vocabularies whose words, possibly partially ordered strings of letters, are routines for sorting, marshalling, transcribing, and so forth, vocabularies. In general they are the hierarchies of programming.

At most this paper is therefore only a programme, with a hint as to how the programme may be achieved. The applications exhibited have been very simple, though not trivial, and with results already known. The mathematics is commonplace and elementary. However, the paper is not intended as a contribution to mathematics, but as a contribution to documentation. Here, especially in retrieval areas, the principle of sympathetic magic, as well as of the unsympathetic variety, is peculiarly powerful. Systems founded on neat and economical theory are neat and economical in other ways.

REFERENCES

1. BRADFORD, S.C., *Documentation,* London; Crosby Lockwood, 1948.
2. BRILLOUIN, L., *Science and Information Theory,* Chap. 4, New York, Academic Press, 1956.
3. FURRY, W.H., and HURWITZ, H., Distribution of Numbers and Distribution of Significant Digits, *Nature 155,* 52/53 (1945).
4. GOUDSMIT, S.A., and FURRY, W.H., Significant Figures of Numbers in Statistical Tables, *Nature 154,* 800/801 (1944).
5. VICKERY, B.V., Bradford's Law of Scattering, *J. Document. 4,* 198/203 (1948).
6. WALL, H.S., *Analytic Theory of Continued Fractions,* Chap. XI, New York, Van Nostrand, 1948.

A Mathematical Theory of Language Symbols in Retrieval[1]

CALVIN N.MOOERS

ABSTRACT. A mathematical model is presented which relates the language symbols of retrieval to the documents retrieved. The model is applied to three families of retrieval systems: those using for language symbols (1) descriptors, (2) characters with hierarchy, and (3) characters with logic. Most information retrieval systems now in use are variations of one of these systems. The similarities and differences between the three systems are displayed by the model. According to the model, a retrieval prescription is represented by a point in a space *P*. This space can be generated by taking the cardinal product of a repertory of simple partially ordered systems. The output of the retrieval system is a subset of documents, and each of these subsets can be represented by a point in a space *L*. The retrieval operation is represented by a transformation from a point in space *P* to a point in space *L*. Two different retrieval transformations are defined. Future elaborations and extensions of the model are outlined.

Most of the mathematical work which has been done toward developing a theory of information retrieval has been directed to studies of machine mathematics or coding mathematics, or to discussions of the gross statistics of retrieval efficiency. Comparatively little work has been done with the theory of the verbal symbols themselves or the way they operate in retrieval. The purpose of this paper is to lay the foundations for a unified mathematical theory for the language symbols of retrieval and to present a mathematical model based on the theory. It will study the algebras or the modes of combination of language symbols, and it will study the way such symbols are related to the subsets of documents which contain the desired information.

In present day writings about retrieval systems and their methodology, there is a confusing variety both in terminology and method with respect to *(a)* the verbal language symbols used, *(b)* the restrictions that are imposed upon the symbols, *(c)* the modes of combination that are permitted between the symbols, and *(d)* the manner in which the symbols or groups of symbols are re

CALVIN N.MOOERS Zator Company, Cambridge, Massachusetts.

[1] Work supported under U.S. Bureau of Standards Contract No. CST-4338.

lated to the documents. Without any unifying theory to cover this variety of terminology and methodology, it is difficult to relate and compare one retrieval system with another, or to relate their individual theories with any of the well-established branches of mathematics or logic whose concepts, results, and theorems would be of assistance to the further development of a general retrieval theory.

This paper presents a mathematical model for the study of retrieval systems. According to this unifying theoretical scheme, it is possible to start from simple postulates and elements and, by systematic procedures, to generate the underlying framework of many of the present retrieval systems. It demonstrates in detail the mode of use of their verbal systems and exhibits the algebra of their symbol combinations and the ways in which the symbol combinations are related to the pertinent subsets in the document collection.

There are a number of important reasons for developing a mathematical model of this kind. It provides a neutral language by means of which the features of a great variety of retrieval systems can be discussed. Since the scheme which is presented can generate the theoretical framework of many different retrieval systems, starting from very simple postulates, it permits easy study of the effect of varying these postulates. We can see how one recognized kind of retrieval system changes into another when even a relatively simple change is made in the elementary combining elements. From this viewpoint, the various retrieval systems are actually special cases generated by the unifying mathematical model and therefore are easily placed in perspective with each other. Because the theoretical scheme described in this paper is based upon well-known mathematical concepts, it has the advantage that it will enable us to relate information retrieval theories to some of the established results from various branches of mathematics and perhaps to extend our theories as a result of that contact.

In this paper, attention is restricted entirely to the verbal symbols, their combinations, and to how they provide retrieval with respect to subsets of documents chosen from a document collection. The paper is not concerned with the details of the retrieval machine or its coding. It presumes that some retrieval machine does exist which is capable of performing the various tasks required by the verbal symbols. We are safe in making this presumption because the general purpose computing machines now available are so versatile and capable that any of the tasks of retrieval described here can be performed on such a machine. For this reason we know that there is at least one class of machine which can do the job. However, the use of these very expensive and complex machines is usually wasteful for performing simple retrieval tasks. Generally, much simpler special purpose machines will be used for retrieval.

PLAN OF ATTACK

A customer desiring to use an information retrieval system actuates it by presenting a "prescription" for the information that he wants. The retrieval system responds to this prescription by indicating to the customer a set of documents from the collection which presumably will furnish the information he desires. In other words, an information retrieval system translates or transforms the customer's prescription into a set of documents.

In all the retrieval systems now in use, the prescription is not given in ordinary English language, but rather in a special language of appropriately chosen retrieval words. These are the "language symbols" with which this paper is concerned. The customer's retrieval prescription is ordinarily never stated in terms of only a single language symbol. It usually is composed of a complex of several symbols taken together. In the same way, the customer is seldom interested in a single document chosen from the collection. His needs are better answered by a small group of documents, that is, a subset of documents chosen from the collection.

For this reason, in studying retrieval systems, our main attention should be focused not upon single symbols nor upon single documents. It should be focused upon the complex of symbols taken together which form a prescription and upon the relationship of this prescription to the subsets of documents that can be chosen from the collection as a whole.

It is helpful to have a convenient way of picturing this relationship (see Fig. 1). The balloon at the left marked *P* represents, in the mathematical sense, a "space" in which each point represents a possible combination of language symbols useful in retrieval. The characteristics of this space will be developed in detail as we proceed. For the moment, we should realize that every possible point in the space represents a retrieval prescription, and conversely that every possible retrieval prescription is represented by a point in this space.

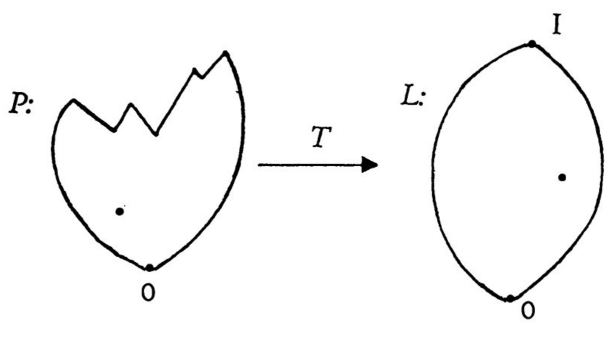


FIGURE 1. The space *P* of all possible retrieval prescriptions, the space *L* of all possible document subsets, and the retrieval transformation *T* associating points of *P* with points of *L*.

The balloon at the right marked *L* is another "space." Each point in space *L* represents a subset of documents that can be formed by choosing documents from the entire collection in a library, and every set of documents that can be formed from the entire collection is represented by a point in this space. This space has a very large number of points. If there are *N* documents in the library, there are 2^N subsets that can be formed on the documents in the collection (including the subset having no documents and the subset having all the documents). For example, if the library has 1,000,000 documents, the number of subsets is $2^{1,000,000}$ or $10^{310,000}$.

There is a point in the space *L* representing any choice of documents that could be taken from the library. A moment's reflection will show that most of the points in *L* represent heterogeneous sets of documents with no unifying similarities in their subject content; such sets are not a useful output to any input retrieval prescription. The sets that are a useful output are a very tiny fraction of the total points in *L*. On the basis of some rather reasonable presumptions about the variety of input prescriptions, I have computed that less than 10^{75} of the points in a typical space *L* for a collection of 1,000,000 would ever be used. Thus, in a relative sense, "almost all" the points in *L* are not useful for output. This fact constitutes the central problem of information retrieval. It is how to find the few useful subsets of documents when almost all the subsets that might be chosen in the collection are not useful.

The purpose of any retrieval system is to receive as an input a prescription which can be represented by a point in *P,* and to convert this input to an output which consists of a citation to, or a set of copies from, some subset of documents taken from the collection represented by a point in *L*. In other words, in terms of the diagram in Fig. 1, if we choose any point in *P* as a prescription, the retrieval system is to emit an output which is represented by some point in *L*. There is a fixed relationship or association between each and every one of the points in *P* to some of the points in *L*. A mathematician calls such an association a "transformation." The retrieval transformation *T* is represented in Fig. 1 symbolically by the arrow directed from *P* to *L*.

If the information in the retrieval system depends upon some machine searching device, or other mechanization, the machine performs the transformation *T*. We should keep clearly in mind that the machine, when it is performing the transformation *T,* is only carrying out the intellectual design of a retrieval system which is represented by the two spaces *P* and *L* and the retrieval association from the one space to the other. If the design of the language symbol aspect of a retrieval system, as in Fig. 1, is ill-conceived, then no machine, however good, can overcome the poor intellectual design. For this reason, we must first pay attention to the way the language symbols operate in retrieval. Only after we are sure that the language symbols perform the way we want them to, should we try to mechanize the system with some machine.

Many details are yet to be filled in before Fig. 1 is really meaningful. The structures of the two spaces have not been worked out, nor has the basis of the transformation been described. I shall begin by considering the space *L*. A mathematician would call this a "finite Boolean algebra." (1, p. 159; 2, p. 318). What this really means is that the relationship between the points of the space *L* are just the relationships that will be found between the various subsets that can be formed on any aggregate of *N* objects, where each subset corresponds to a point of the space. Given a set of *N* objects, these are some of the operations and relationships: we can form subsets of the objects; we can combine or add any two of the subsets; we can take the intersection or the common members of any two of the subsets; and, upon comparing their membership, we can find that some of the subsets are included within others. Since the number *N* of objects is finite, the number of subsets is finite and thus we have a finite algebra. When we have a familiarity with these elementary operations upon subsets, we know some of the most important properties of a finite Boolean algebra and thus of the space *L* (cf. 2, p. 311).

Space *P* of the retrieval prescriptions is much more complicated than *L,* and we shall have to study it in more detail. Part of its complication results from the many differences between one retrieval system and another in the basic methodology of language symbols. However, it is difficult to study differences in methodology if we confine our attention only to the large and complicated space *P*. The problem becomes much simpler if we have some way to decompose *P* into simpler components. Fortunately, there is an easy way to do this.

We start with the repertory *R* consisting of a finite number of simple structures, as indicated in Fig. 2. Each of these structures, represented by one of the little diagrams enclosed in the brackets, stands for a simple retrieval element such as a descriptor or some other kind of term which I shall call a "character." These simple structures can be combined or multiplied together to give a

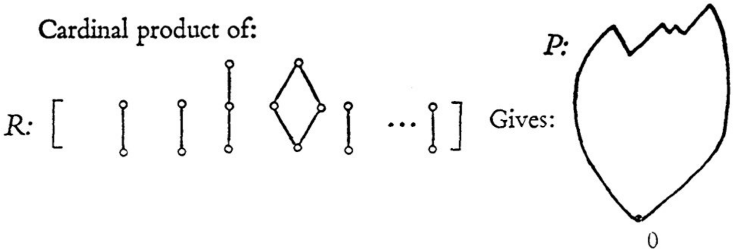


FIGURE 2. The repertory *R* of simple structures whose cardinal product is the space *P* of all possible retrieval prescriptions.

"cardinal product" or a "direct product." The cardinal product is also a structure, but it is more complicated than any of its factors. By combining the simple structures by the cardinal product, we actually can generate the space *P* with all its properties. Therefore, in order to get the answers to many of our questions about *P,* we can sometimes do it more easily by going back to the repertory *R* with its simple, easily understood elements.

This viewpoint has a number of advantages. We can see exactly how space *P* is built up or generated. We can more easily see the effect of employing different kinds of characters in the repertory. We can better understand exactly what we do when we make a prescription for retrieval or when we analyze a document to put it into the system. By a better understanding of *P,* we can also better understand the nature of the retrieval transformation *T* between the space *P* and space *L*.

However, before we can continue our investigation into the detailed features of retrieval systems, it will be necessary to introduce a number of mathematical tools and concepts that we will need in the exposition to follow.

CONCEPTS EMPLOYED

The first concept concerns the language elements which prescribe retrieval. In this paper I shall be primarily concerned with the class of language elements which I have called "characters." I define a character as a verbal symbol which *(a)* can be independently manipulated in prescribing a retrieval selection, *(b)* does not decompose into two or more such independent symbols, *(c)* has a definite meaning or interpretation, and *(d)* comes from a finite repertory.

A string of Arabic numerals making up a decimal classification number, e.g., 512.3, is a character: it is complete in itself; it cannot be broken into two or more parts; it has a definite meaning; and it is taken from a repertory which is finite and not rapidly changing.

A descriptor (3) is a character because it clearly satisfies each of the four required points: it can be independently manipulated; it cannot be split up; it is precisely defined by a scope note; and it comes from a finite and actually quite small repertory.

A subject heading may or may not be a character, depending upon the care with which it is defined and used. Sometimes subject headings are used without scope notes or any other careful definition of meaning. Sometimes they are permitted to be generated ad lib without any particular limitation or control. Whenever these situations prevail, subject headings are defective characters.

"Uniterms" as introduced by Taube (4) have been used in a great variety of ways and without very much standardization. Since they often tend to be words merely taken from text, they are likely to be defective as characters because they have no controlled meaning. The typical repertory of Uniterms is also indefinite and "open-ended," and it is questionable if such a repertory can be called "finite" in the sense that any finite upper bound can be put on the total number of Uniterms for a system. If Uniterms were made more precise, by using scope notes, and if their number and their proliferation were controlled, they would fall within the definition of characters.

The next concept is that of a "partially ordered system" (1, p. 1; 2, p. 326). A very simple example of a partially ordered system is the hierarchy of people working in an office. The partial ordering relation is based upon who may receive orders from whom. If two people in the office are from different departments, neither may receive orders from the other. However, if they are both in the same "line of authority," then one of them must take orders from the other. An organization chart is actually a diagram of this partially ordered system. It shows who takes orders from whom. Note that this is called a "partially ordered," rather than a "totally ordered," system in order specifically to indicate that the ordering relation is true of *only certain pairs* of people, and not of all pairs that can be formed.

More abstractly, a partially ordered system is a set of elements together with a specific ordering relationship such that, given any two elements x and y from the set, it is possible to say of them that x precedes y, or y precedes x, or that neither precedes the other. When x precedes y, it is signified by $x \leqq y$. (Where Birkhoff uses "included," I have used the word "precedes" because of the possible confusion that might arise over the possible meanings of inclusion in discussing some other problems of retrieval.)

As an example of a partially ordered system in retrieval, we can take the points in the space L. Each of these points represents a subset of documents. If the partially ordered relationship is that of "being included within," some of the subsets are included within some of the others. For other pairs of subsets, neither subset is completely included within the other, and therefore neither precedes the other in the partially ordered system.

A partially ordered system can be diagramed. This has been done in Fig. 3 for several simple systems. The individual elements of the partially ordered system are represented by small circles. The partial orderings are shown by the connecting lines. If element x precedes element y, x is below y, and it is possible to trace a continuously descending path with no up-and-down zigzags from element y down to element x. The path from element y to element x may pass through several other elements, or x and y may be adjacent. No horizontal lines are allowed in a diagram of this kind.

Some partially ordered systems have a "greatest element," or a "least element," or both. Referring to Fig. 3, element 5 of system *(A)* or element 8 of *(D)* are the greatest elements of these partially ordered systems. In the same fashion, element 1 of *(A)* or element 4 of *(C)* are least elements of their systems. By convention, the greatest element of a partially ordered system is usually denoted by I, and the least element by 0. Not all partially ordered systems have a greatest or least element. The partially ordered system *(C)* has a least element, but it has no unique greatest element.

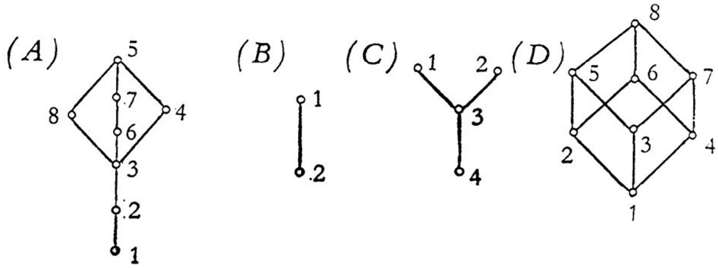


FIGURE 3. Four examples of simple partially ordered systems.

The concept of "level" is useful in the discussion of partially ordered systems (cf. 1, p. 11); however, Birkhoff uses the terms "dimension" or "height." Referring to Fig. 3, the partially ordered system *(C)* has two distinguishable levels above the bottom element. To determine the level number for an element, we give the least element in the partially ordered system level number 0, and we count the number of elements as we move up in the diagram to the element whose level number is desired. Thus in the partially ordered system *(D),* element 5 is at level 2 and element 8 is at level 3. In many partially ordered systems, however, the concept of level does not have any meaning. If the system has no least element, or if there are several paths up to the given element, along different paths giving different counts, then the concept of level cannot be employed. For example, in the system *(A)* no consistent level can be assigned to the point 5. Sometimes it is useful to speak of levels counting down from the top, and this is done in a similar fashion but with the directions reversed.

The concept of cardinal or direct product between partially ordered systems plays a central role in the systematic method of generating the basic structures of retrieval systems. Given two partially ordered systems X and Y, whose typical elements are denoted by x and y, the cardinal product XY is also a partially ordered system. It is composed of all couples of the form (x,y), and it is partially ordered by the rule that (x,y) precedes (x',y') whenever x precedes x' in X, and y precedes y' in Y. The formation of the cardinal product can be diagramed. In Fig. 4 are shown two partially ordered systems along with their cardinal product. The individual elements of each of the original systems are numbered so that it is possible to see how the cardinal product partially ordered system is formed from the couples, and how the ordering relation follows from the definition. This cardinal product partially ordered system of Fig. 4 is very simple, since it is formed from systems having relatively few elements. When the starting systems are more complicated, the cardinal product becomes very complex.

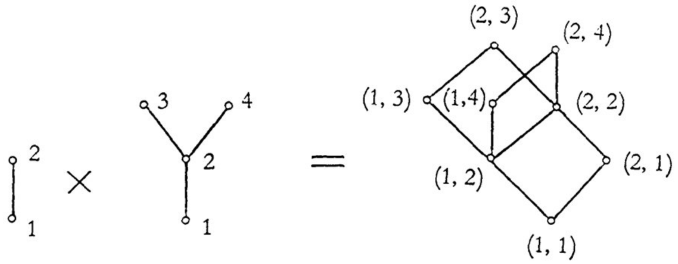


FIGURE 4. Illustration of the formation of the cardinal product of the two systems shown at the left.

The cardinal product of three or more partially ordered systems can also be formed. We merely multiply two of the partially ordered systems together, getting a product, and then multiply the next system into the product. This is permissible because the cardinal product is associative. Thus we can take the cardinal product of all the partially ordered systems in a repertory *R,* as in Fig. 2. As we shall see, this cardinal product system actually is the partially ordered system *P,* whose elements constitute the space in which retrieval systems are formulated.

According to another useful concept, we consider any two elements x and y in a partially ordered system, and we wish to identify the largest element z which precedes both x and y. In terms of the diagram for a partially ordered system, z would be the highest element in the diagram from which one can draw ascending lines to both of the elements x and y. Such an element z is a lower bound to both x and y; in fact, it is a "greatest lower bound" or "g.l.b." The element z is also called the "cap" of the two elements, and this is denoted by $x \cap y = z$. In Fig. 3 *(D)* the cap of elements 5 and 7 is the element 3.

In the same manner we can speak of upper bounds. We define the "least upper bound" or "l.u.b." w of two elements x and y in a partially ordered system as the smallest element which is preceded by both x and y. The least upper bound w of two elements is often called their "cup," and this is denoted by $x \cup y = w$. The cup of elements 5 and 4 in *(D)* of Fig. 3 is the element 8. It should be noted that cup and cap do not always exist in partially ordered systems. In system *(C),* the cap of element 1 and 2 exists, but not their cup. In many situations, particularly when dealing with subsets formed upon an ag gregate of objects, cap is associated with the logical product or the intersecting part of the subsets.

We are now able to introduce the concept of a "lattice." (1, p. 16; 2, p. 328). Fairthorne was perhaps the first to point out the value of lattices for gaining an insight into retrieval systems (5, 6). A lattice is defined as a partially ordered system in which *every* pair of elements from the system have *both* a cup and a cap. For example, the partially ordered systems *(A), (B),* and *(D)* of Fig. 3 are all lattices. The system *(C)* is not a lattice because cup does not exist for all pairs of elements.

One of the most important examples of a lattice is the partially ordered system composed of all the subsets formed upon a finite aggregate of *N* objects. Thus the document subsets that can be formed from a library collection form a lattice. Actually, it is a very special kind of a lattice, and it is called a Boolean algebra or Boolean lattice. Since the subsets of the documents do form a lattice, the space of these subsets in Fig. 1 is denoted by *L*. However, the document prescriptions do not always form a lattice, although they do usually form a partially ordered system. For this reason the space of document prescriptions is denoted by *P*.

We now have the main conceptual tools which will enable us to study the theory of a variety of information retrieval systems.

RETRIEVAL SYSTEMS BASED UPON DESCRIPTORS

Descriptors are perhaps the simplest form of characters, and thus the kind of retrieval system based upon them is simpler than any of the others. The following are the important facts about descriptor retrieval systems (3). When a document is analyzed to be entered into a descriptor retrieval system, the message of the document is tested against each of the descriptors in the repertory. The descriptors whose meanings provide retrieval clues to the message of the document are then associated with the document. Thus for each document in the collection, there is a delineating subset of descriptors. A retrieval prescription is given also in terms of a subset of descriptors. When the retrieval operation is performed, the prescribing subset will be included within some of the document subsets, and it will not be included within some of the others. This provides the criterion of selection. The output of the retrieval system consists of citations to those documents whose delineating subsets include the prescribing descriptor subset. Thus a document will be retrieved if the delineating subset of the document contains each and every one of the descriptors in the prescribing subset. If the document subset has additional descriptors, the document is still retrieved.

Let us now examine the descriptor retrieval system from the standpoint of the mathematical model sketched in Fig. 1. The space *L* of the document sets is a Boolean algebra typified by the subsets formed on a finite aggregate. It poses no problems. The space *P* of the retrieval prescriptions does demand our attention.

The space *P* is the cardinal product of all the two-element partially ordered systems in the repertory, where each of these partially ordered systems represents the behavior of a single descriptor. This is diagramed in Fig. 5. The partially ordered system for an individual descriptor is diagramed in Fig. 6. Let a hypothetical descriptor, which might have the meaning of "red" or "airplane" or "low temperature," be designated by *A*. In either delineation of the document during an analysis, or prescription of it for retrieval, a descriptor can behave in only two ways. On the one hand, the descriptor *A* may be asserted about the document as providing a clue to the message in the document. When this is the case, we represent it by the upper point in the diagram of Fig. 6. The other possibility is that no assertion one way or the other is made about the descriptor *A*. This situation is represented by the lower point marked 0 in the diagram of Fig. 6. These two possibilities, represented by the two points, are partially ordered by the partial ordering relationship: "being included within as a special case." The assertion of *A* is certainly "included within as a special case" of 0, which is the lack of any assertion. This is the sense in which any descriptor and its behavior are quite well represented by this kind of simple, two-element partial ordered system.

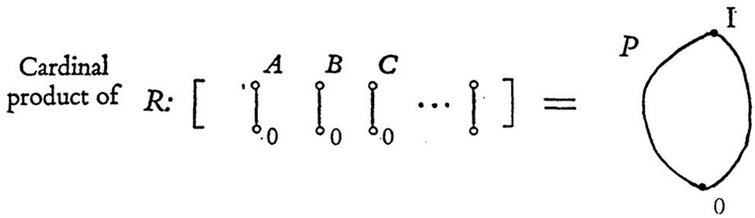


FIGURE 5. For descriptor systems, the partially ordered system *P* is the cardinal product of simple two-element partially ordered systems.

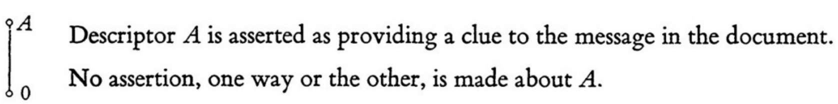


FIGURE 6. The partially ordered system for a single descriptor.

The formation of the cardinal product for the partially ordered systems of several descriptors is shown in Fig. 7. Here the repertory has three descriptors. The complete cardinal product is shown at the right in the figure, and the meaning of each element is written out. For example, the element marked *AB* means that both descriptors *A* and *B* are asserted. This figure also shows the way the notation is related to the definition of the couples which define the cardinal product. For example, taking element *A* in the cardinal product we have: ((*A*,0),0)=(*A*,0,0)=*A*. A similar relationship holds for the other elements in the cardinal product system.

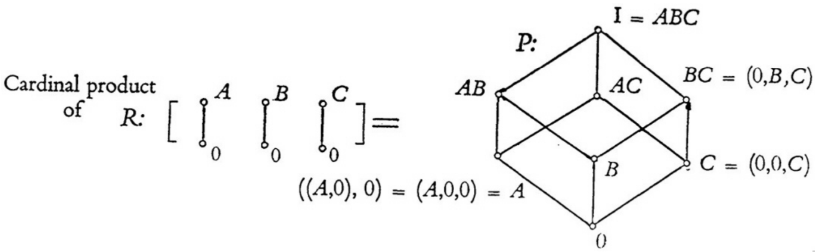


FIGURE 7. This shows the formation of the cardinal product for three descriptors. Each of these partially ordered systems is a Boolean lattice.

The space *P* of the retrieval prescriptions for a retrieval system based upon descriptors is a Boolean lattice. The easiest way to see this is to observe that the system *P* can also be built up by forming all of the possible subsets of the descriptors in the repertory. As has been mentioned before, any partially ordered system of all the subsets of a finite aggregate is a Boolean lattice.

It should be observed that the assertion of any subset of descriptors from the repertory is equivalent, for these descriptors, to taking the upper points in Fig. 5, and for the descriptors which are not asserted, to taking the lower points. Thus, in a way, we take an action upon each of the descriptors in the repertory. As is illustrated in Fig. 7, any such action asserting a subset of descriptors from the repertory is equivalent to asserting a single point in the lattice *P*. This demonstrates, for retrieval systems based upon descriptors, that every delineation of a document in terms of a descriptor set, or every prescription for retrieval in terms of a descriptor set, is represented by a single point in the space *P*.

We can now describe the retrieval transformations from points in the space *P* to the points in the space *L*. The first transformation T_1 is the easiest to describe, but it is not the most useful in actual retrieval systems. Although it is not used at all in descriptor systems, I mention it here because it will be used for comparison purposes. Each document in the collection has a subset of descriptors assigned to it, i.e., each document is associated with some particular point of *P*. Looking now to a typical document subset, such as represented by any point in the space *L*, it is evident that the individual documents in such a subset will usually be associated with a wide variety of points in the space *P*. There will be some points in *L* which are unusual in that they represent the largest set of documents such that all the documents associated with the set shown by the point will have the same set of descriptors. In other words, all the documents of such a point of *L* will have descriptors corresponding to the same point of *P*. Each of these points in *P* therefore has a unique association with a corresponding point in *L* (different from the 0 element). On the other hand, many of the combinations of descriptors, represented by other points of *P,* will not be assigned to any of the documents in the collection. Such points correspond to the set having no documents whatsoever, that is, such points of *P* correspond to the null set of documents represented by the point 0 in the diagram of *L*. Thus it is possible to associate every point of *P* with some point of *L,* though not every point in *L* is associated with a point of *P*. This association between the points of *P* and some of the points of *L* provides the definition of the transformation T_1. Given any point in *P,* the transformation T_1 then names a point in *L*.

In considering the retrieval properties of the transformation T_1, we see that when we make a retrieval prescription by naming a single point in *P,* the transformation T_1 gives us a subset of documents all which possess exactly the set of descriptors corresponding to the retrieval prescription point. For many purposes, this is not the best kind of selection for retrieval. The difficulty with this kind of selection is that when we name a set of descriptors to prescribe a search, we are interested not only in the documents which have exactly the prescribed set of descriptors, but also in any other documents having any additional descriptors which may not be specified. The transformation T_1, when used for selective purposes, will exclude all such documents having descriptors other than those in the descriptor subset corresponding to the prescribing point of *P*. To remedy this defect, we must go to a slightly more complicated transformation, which I have designated T_2.

The transformation T_2 is the basis of selection in actual descriptor retrieval systems. Although the transformation T_2 has already been described at the beginning of this section, we now want to describe it in terms of the points in spaces *P* and *L*. Given any point x in the space *P,* there is a large family *X* of other points in *P* which are preceded by the point x. As shown in Fig. 8, all these points lie above the point x in the partial ordering diagram. Looking now to the document collection, there are many documents whose assigned subset of descriptors is one of the points in the family *X*. We can then consider the largest set of documents such that each document in the set has a subset of descriptors represented by some point in *X*. Let this set of documents be represented by a point $x^\star$ in *L*. We can now define the point $x^\star$ in *L* to be the transformation according to T_2 of the point x in *P*.

From this definition of T_2, each of the points in *P* has a corresponding image point in *L*. On the other hand, not all the points of *L* are images of such a transform, in fact, only a very few of them are. It is useful to designate the points of L which are the images of the points in P, through the transformation T_2, by the designation $L^{\star}$. Then every point in P is associated by T_2 with some point in $L^{\star}$, and every point in $L^{\star}$ is associated (through the inverse of the transformation T_2) with some point in P. If we were to give a characterization of the partially ordered system $L^{\star}$, we could well call it the "retrieval skeleton of the system of document sets of the collection." This name is appropriate because the system $L^{\star}$, although containing only a very small part of the system L, contains all the points of L which are of primary interest for the retrieval.

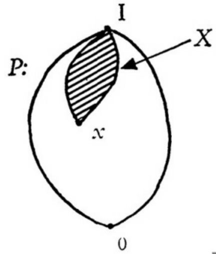


FIGURE 8. The class of points X in the system P consists of all the points which are preceded by the prescribing point x.

It is enlightening to make some comparisons between the transformations T_1 and T_2. One interesting comparison is based upon observing the transform of an ascending path drawn through space P. (Refer to Fig. 9.) By an ascending path, is meant a path drawn from 0 at the bottom, through one point after

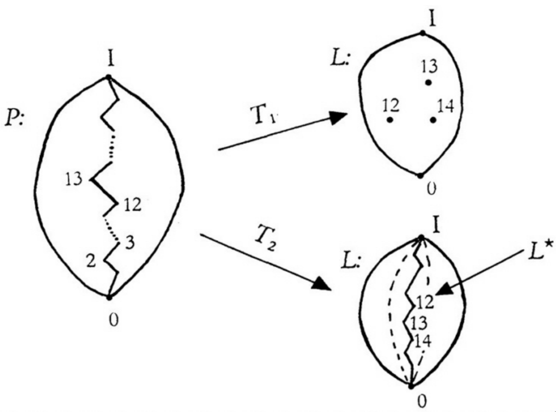


FIGURE 9. A continuously ascending line drawn in P transforms by T_1 into points in L representing non-overlapping subsets. The transformation of the line by T_2 is a descending line in L. The lattice $L^{\star}$ is the set of all the image points of P by the transformation T_2.

another, going continuously upward through a selection of points, until the topmost point I is reached. Such a path represents in the space *P* what happens when we first start out with no descriptors at all in a prescription, then take one descriptor, then add another descriptor to it, then add a third descriptor to those we have, and so on until our accumulation of descriptors finally includes all the descriptors in the repertory, bringing us to the point I. When the accumulation has *k* descriptors in it, the set of descriptors is represented in *P* by a point at level *k* in the diagram.

If we apply the transformation T_1 to the individual points along such a path, this is what happens. For the first several descriptors (*k* small), all the points transform into point 0 of *L* because most documents will always have more than only one, two, or three descriptors. Next, there will be some first point, such as at level 12, for which there are some documents having the corresponding set of descriptors. The point in *L* for this document set is at some point in the interior of *L*. There may also be documents corresponding to the next point on the line, at level 13. The transform of this next point will also be somewhere in the interior of space *L*. However, these two points will not be ordered with respect to each other in space *L* because they represent mutually exclusive sets of documents. Neither of these image points can precede the other according to the partial ordering relationship of set inclusion that prevails in the system *L*.

Thus we have the interesting situation in which the point at level 12 does precede the point at level 13 in space *P,* but the transforms of these points according to T_1 are points which have no ordering relationship in *L*.

At a sufficiently high level in *P* along the ascending line, that is, for sets of a sufficiently large number of descriptors, there will be no corresponding documents in the collection. For these points in *P,* the transform according to T_1 will again be the point 0 in *L*.

Thus the points along a continuously ascending line in *P* transform according to T_1 as follows: first, into the point 0 of *L,* then into a scattering of (non-ordered) points in the interior of *L,* and finally, again into the point 0 of *L*.

This behavior may be contrasted with the transform according to T_2 of the same points along the line. Referring again to Fig. 9, the point 0 of *P* goes over into I of *L*. This is because a prescription involving no descriptors will allow all the documents in the collection to be selected. Conversely, the point I of *P* goes over into the point 0 of *L,* because none of the documents in the collection will be associated with all the descriptors. Moreover, the transform of each of the points on the ascending line drawn in *P* is some point in *L* lying between I and 0 (or including I and 0). The points near the top of the line in *P* all go into the 0 element of *L,* since there are no documents having this many descriptors. However, the transform of the first point above the 0 element *P* is not a point I in *L,* because the restriction imposed even by a single descriptor will prevent the entire collection (represented by I) from being selected. As we go up from point to point along the line in *P,* the transforms of the points are located in descending positions in *L* (and are in adjacent descending positions in $L^\star$). Moreover, the image points in *L* of the points of the line in *P* are all ordered with respect to each other. (None of the sets are mutually exclusive.) Where the point at level 12 in *P* precedes the point at level 13, the image point in *L* of the point 13 precedes that of point 12. Note that the direction of the partial ordering is inverted by the transformation T_2.

In terms of the practical operation of an information retrieval system, what I have diagramed is the well-known fact that as one adds more and more descriptors to a retrieval prescription, the set of retrieved documents becomes smaller and smaller, and that each of the smaller sets of documents is included within the larger set which is obtained with fewer descriptors in a prescription.

It should be noted that some of the important machine systems and their codes for operation according to the descriptor retrieval system do not reproduce the transformation T_2 exactly. For instance, the Zatocoding system, which uses code patterns of a random nature and which superimposes these patterns into the coding matrix, is one of these coding systems (3, 7). Such a method of coding will produce only a close approximation to the transformation T_2. By suitable design of the coding any degree of accuracy in reproducing T_2 can be achieved. All the desirable properties of the transformation T_2 are preserved except for the addition of a negligibly small amount of "noise" which can be controlled. However, a discussion of machines and their coding systems is beyond the scope of this paper.

The points of the space *P* form a lattice, and therefore the operations cup and cap are always defined. For any points *x* and *y,* $x \cup y = z$. However, by T_2, each of the points *x, y,* and *z* has its transform point in *L*. Where the transform of *x* is indicated by $x^\star$ (and so on, for the other elements), we have, in the lattice $L^\star$, the relationship $x^\star \cap y^\star = z^\star$. However, this relationship *does not* hold in the lattice *L*.

Similarly, for any three elements in *P* related by the cap operation, their corresponding transform points in $L^\star$ are related by the cup operation. Again, the cup relationship between the transformed points does not hold in *L*. These results may be compared to those of the transformation T_1, for which neither the cup nor the cap operation on elements in *P* has any valid relationship to the cup and cap operations in *L*.

From these remarks we see that algebraic operations involving cup and cap operations in *P* are valid in $L^\star$, although the cup and cap are interchanged;

while these operations in P have no corresponding operations in L according to either T_1 or T_2. In practical retrieval system operation, the preservation of the operations cup and cap through the transformation T_2 into the lattice $L^{\star}$ mirrors the way in which the descriptor combinations are related to the selected sets of documents. With these observations, we go on to the next type of retrieval system considered.

RETRIEVAL SYSTEMS BASED UPON CHARACTERS WITH HIERARCHY

The language symbols of retrieval employed in a hierarchical classification system are called "characters with hierarchy." There are a number of examples of retrieval systems employing such characters. Very familiar are the Library of Congress and the Dewey Decimal classification systems. Slightly different in basic organization is the U.S. Patent Office classification system. Also falling within this category are the characters in the technique of interlocking sets of descriptors for the representation of structure (8). The Colon Classification of Ranganathan also uses characters with hierarchy (9).

There is a certain difficulty in studying characters with hierarchy because some of the systems have become so familiar to us with long use that we ordinarily do not think about their basic structure. Thus, although several of the examples named may seem to be the same in many respects, they have quite different postulates. One of the purposes of this section is to display which postulates are similar and which are different for the systems.

As an example of the behavior of characters with hierarchy, consider the two retrieval characters "shoes" and "clothing." Because shoes are a kind of clothing, these two retrieval characters are not independent. When we are talking about shoes, we mean a specific kind of clothing. Also, when we are talking about clothing, we often intend to include shoes within the compass of our discussion.

These two characters do not behave like descriptors because, when we take the two of them together, i.e., "shoes—clothing," one of them loses its purpose in any retrieval operation. If the two characters taken together are intended to cover all the ground mentioned by either of them, then their meaning is covered by the single character "clothing." On the other hand, if the two characters taken together are intended to refine and to narrow the meaning of each other, then their meaning is completely covered by "shoes." Ordinarily the second of these two possibilities is intended, and we shall make that interpretation in the discussion which follows.

Since some characters with hierarchy absorb others, they evidently are partially ordered. The diagram of the partially ordered system for the charac ters "shoes" and "clothing" is shown in Fig. 10. In explanation of this diagram, consider the delineation of the documents as they are analyzed to be placed into the retrieval system. If we make no assertion whatsoever about shoes or clothing, this is represented by the null element 0. The simplest factual assertion that can be made in delineation is "clothing," represented by *(a)*. A more exact assertion is "shoes," but this character, during delineation of the document, must also carry along with it the meaning "clothing." For this reason, "shoes" is considered to be a refinement of "clothing," and is represented by *(ab)*.

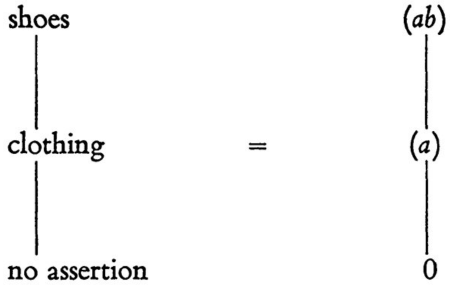


FIGURE 10. The partially ordered system based upon two characters with hierarchy.

In the notation which I have adopted, the letters *a* and *b* without the parentheses around them are not characters. I shall call such letters "simple elements" or merely "elements." The purpose of the elements is to represent the logical structure of characters with hierarchy. The elements cannot be used independently for retrieval operations; only characters can be so used. To indicate when an element, or a string of elements, becomes a character, I shall enclose the letters or string of letters in parentheses. Although we might say that element *b* of Fig. 10 has the meaning "shoes—irrespective of clothing," this kind of analysis has not generally been used in hierarchical classification systems. I shall avoid it here. Only characters will have meanings usable in retrieval.

According to the logical structure of a hierarchical classification system, when we delineate a document with the character "shoes," we also intend the assertion to include the assertion "clothing." In another example, if a book or document is delineated by the character 512.3 from the Dewey Decimal Classification system, we also mean that this same document is delineated by the characters 512.–, 51–.–, and 5– –.–. In general, if the elemental representation of a typical character is *(abcd),* we intend that any document delineated by this character is also, by implication, delineated by any of the characters which can be formed by successively crossing off elements from the right-hand end of the representation. In another way of looking at this matter, we refer to the partially ordered space *P* of Fig. 11, in which each point of the space represents a possible delineation of the document. Then, if the document is delineated by a character at point *x,* by implication, it is also delineated by all the points preceding *x,* as shown by the shaded area of Fig. 11.

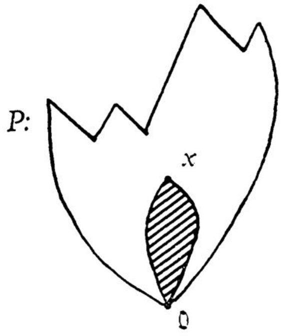


FIGURE 11. In a system *P* of characters with hierarchy, if a document is delineated by a point *x,* it is by implication delineated by all the points preceding *x,* as shown in the shaded area.

This brings out an interesting dual aspect of a hierarchical classification system. On the one hand, the characters are ordered in a hierarchy, and sometimes this ordering of characters alone seems to be an end in itself. On the other hand, the delineation of a document by one character implies its simultaneous delineation by all the characters that precede it in the hierarchical scheme. This has practical retrieval consequences. However, we should remember that this multiple delineation is only implicit. The degree to which it is actually made use of for practical retrieval purposes differs greatly from system to system.

The cardinal product can generate the retrieval language system for characters with hierarchy. This occurs when the cardinal product of a repertory of simple partially ordered systems is formed. In the case of characters with hierarchy, the simple partially ordered systems of the repertory are such as those shown in Fig. 10. Each partially ordered system of the repertory is a chain; that is, it is a single line of characters, with the most particular one at the top, descending down through several levels to the 0 element. In Fig. 12 at the left are shown two such chains. Their cardinal product is shown in the center. However, this cardinal product must be considerably simplified before it is really meaningful. For instance, the character *(a,a)* might be crudely interpreted as "clothing-clothing." This is nonsense. What is really meant is merely "clothing." Therefore such symbols as *(a,a)* must be reduced to *(a)*. In the same way, the other symbols in the cardinal product must be simplified as follows. The commas are dropped out. Letters which repeat other letters are eliminated. If there are any letters other than 0, the 0 is eliminated. The remaining letters are put in alphabetical order. The result is the elemental representation of the character. This simplification applied to the center diagram produces the upper right-hand diagram of Fig. 12.

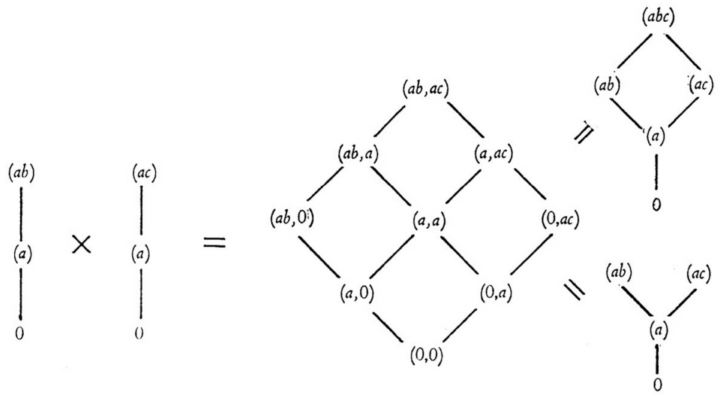


FIGURE 12. The cardinal product of the two partially ordered systems shown at the left has been taken, the intermediate product is shown at the center, and after the appropriate reductions have been performed, the final products are shown at the right. At top right is the system with weak hierarchy, at lower right is the system with strong hierarchy.

The new character *(abc),* which has appeared as a result of the cardinal product, needs some interpretation. If character *(ab)* has the same meaning as in Fig. 10, and if character *(ac)* is given the meaning "trousers," then we can interpret the character *(abc)* as meaning "shoes and trousers integral in one garment of clothing."

There are systems of hierarchical characters which do not permit the formation of systems of the type shown in the upper right-hand diagram of Fig. 12. In general, these systems do not permit a character to be preceded immediately by two other characters. For such systems, the manner of reduction must be supplemented so that the cardinal product of the original character systems is reduced to the form as shown in the diagram at the lower right-hand corner of Fig. 12.

The two situations illustrated at the right in Fig. 12 must be distinguished. Systems which do permit some characters to be preceded by more than one character, I call "systems with weak hierarchy." The system of classification used in the U.S. Patent Office has weak hierarchy. Systems which do not permit any character to be preceded immediately by more than one character, I call "systems with strong hierarchy." The Dewey Decimal Classification has strong hierarchy. The diagram of a strong hierarchical system is called a "tree" because none of the lines ever rejoins as you trace upwards in the diagram. In my opinion, systems with weak hierarchy tend to be more versatile in retrieval.

We shall now consider postulational treatment of characters with hierarchy, based upon properties assigned to the elements.

1. There is an ordered sequence (alphabet) of simple elements, e.g., *a, b, c,*· · ·, etc.

2. Each character is represented by a string of simple elements taken from the sequence, and the elements in the string are arranged in the same order as in the sequence.

3. Strong hierarchy: In a system of characters with a strong hierarchy, each element in the sequence has a unique predecessor element in the sequence. When any element occurs in any string, it is always immediately preceded by its predecessor element. Elements which have the null element as a predecessor may begin a string.

4. Weak hierarchy: In a system of characters with weak hierarchy, each element in the sequence has one or more predecessor elements (including the null element), and it may also have one or more co-predecessor elements. When an element occurs in any string, it must always be preceded immediately by all its predecessor elements except when a co-predecessor element is in the string, in which case the co-predecessors may come between the element and its predecessors.

The operation of these postulates for characters with strong hierarchy is shown by example in the table of Fig. 13. Typical strings of elements forming characters are shown. Also listed are the elements and their predecessors. We notice that strings such as *(ac)* are not allowed because element *c* must always be preceded by element *b*.

Element	*Predecessor*	*Character formed*
a	Null	*(a)*
b	*a*	*(ab)*
c	*b*	*(abc)*
d	*b*	*(abd)*
e	*d*	*(abde)*

FIGURE 13. Illustration of the application of the postulates to characters with strong hierarchy. Note that the combinations *(ac), (ad), (abcd),* or *(ace)* cannot represent characters.

The application of these postulates to the formation of characters with weak hierarchy is shown by example in the table of Fig. 14. We notice that element *e* must always be preceded by *b, c,* and *d*. Element *d* must be preceded by *a,* and *d* may or may not be preceded by *b* and *c*. Thus, characters formed according to weak hierarchy have a great deal more variety, because of the fewer restrictions, than characters with strong hierarchy.

From these examples, and by the postulates, we can see that the system of characters in the upper right-hand corner of Fig. 12 is compatible with weak hierarchy, while the system in the lower right-hand corner is compatible with strong hierarchy.

Element	*Predecessors*	*Co-predecessors*	*Character formed*
a	Null		*(a)*
b	*a*		*(ab)*
c	*a*		*(ac)*
d	*a*	*b, c*	*(abd), (acd), (abcd), (ad)*
e	*b, c, d*		*(abcde)*

FIGURE 14. Illustration of the application of the postulates to characters with weak hierarchy. Note that combinations *(ade), (bc),* or *(ae)* are among those that cannot represent characters.

In any system of characters with weak hierarchy, the system becomes more and more versatile as the different elements are allowed to have more and more co-predecessors, and also as the requirements for predecessors are removed. If this process is carried out to the limit, we have a situation in which the elements have no required predecessors (other than the null element), and they may have any other element as co-predecessors. When this happens, we can form characters by selecting any set of elements from the alphabet and listing them in order to form a string making up a character. With such a string, we can form a new character by deleting any element from the string. A moment's reflection will show that, when the postulates are relaxed to this extent, we have a system of characters which is almost identical with the descriptor system. The elements behave as descriptors, and the strings of elements forming characters correspond to sets of descriptors.

This comparison provides an explicit demonstration of how our mathematical model can relate, through use of appropriate common terminology, two different kinds of retrieval systems which are ordinarily thought to be quite different.

In order to generate the partially ordered system of characters *P,* we must take the cardinal product of the set of partially ordered systems in the repertory. As yet, these partially ordered systems of the repertory have not been defined. Therefore, we require the following additional postulates in order to define the systems making up the repertory *R*.

5. The partially ordered systems of characters in the repertory are chains. The top character in each chain is called the “dominant character,” and it is represented by a string of elements.

6. For characters with strong hierarchy, there exists a dominant character

for each element which is not a predecessor element to any other element. This dominant character is represented by a string of elements ending in the element which is not a predecessor. (Thus there are as many dominant characters as there are elements which are not predecessors.) Each of the characters in the chain preceding a dominant character is formed by deleting one after another of the elements at the right end of the string of elements which designates the dominant character.

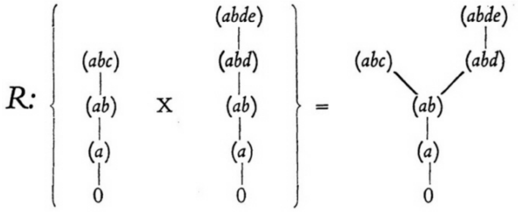
7a. For characters with weak hierarchy, there is a dominant character for each element which is not a predecessor of some other element, and such a dominant character is represented by a string of elements beginning at the right with the non-predecessor element and preceded by its predecessor elements only, the co-predecessor elements being systematically omitted.

7b. The characters in a chain which precede each dominant character are represented by strings in which successive elements are crossed off one after the other from the right-hand end. As this crossing off proceeds, certain elements in the string which were predecessors suddenly convert to co-predecessors. When this happens, such elements are eliminated as a group.

7c. A dominant character is then formed for each of these converted co-predecessors, with the co-predecessor element beginning the right end of a string which is built up of predecessor elements again. The characters in the chain preceding such new dominant characters are again formed by crossing off from the right.

7d. This process is continued until all of the dominant characters and their strings have been formed.

In order to illustrate the principles of postulates 1, 2, 3, 5, and 6, the system of characters with strong hierarchy described by the table in Fig. 13 has been diagramed in Fig. 15. To the left is the repertory consisting of two chains, each headed by a dominant character. The cardinal product of these chains, after all the reductions, is shown by the diagram in the center. Since we are dealing with strong hierarchy, none of the paths of such a diagram, once they

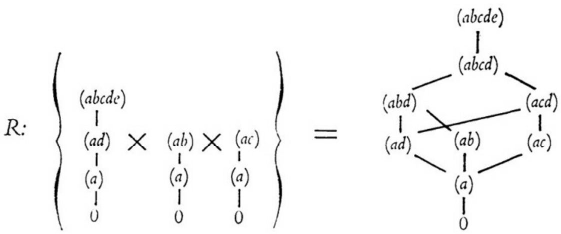


Element	Predecessors
a	0
b	a
c	b
d	b
e	d

FIGURE 15. Diagram illustrating the cardinal product and reductions according to postulates for strong hierarchy.

are separated, will ever join again as we go upwards. The diagram is a tree. At the right in Fig. 15 is a table giving the predecessors to each of the elements.

The system of characters with weak hierarchy from the table in Fig. 14 is diagramed in Fig. 16. At the left are three chains, each headed by a dominant character, which are found according to postulate 7. The cardinal product of these three chains, after all the reductions, is shown at the right. The table giving the predecessors and co-predecessors for each element is at the bottom. Careful study of Fig. 16 will demonstrate the application of postulates 1, 2, 4, 5, and 7.



Element	*Predecessors*	*Co-Predecessors*
a	0	
b	*a*	
c	*a*	
d	*a*	*b,c*
e	*b,c,d*	

FIGURE 16. Diagram illustrating the cardinal product and reductions according to postulates for weak hierarchy. (*Note.* Although the same letters are used to represent the elements as in the system shown in FIG. 15, no relationship between the two systems is intended.)

At this point, I wish to mention the Colon Classification developed by S.R. Ranganathan of India (9). This ingenious classification provides somewhat greater freedom than the conventional methods of decimal classification. It can be described in terms of our terminology and postulates as a hierarchical system of characters in which the repertory of chains of characters is partitioned into several sub-repertories. Within any sub-repertory, the characters have strong hierarchy. However, between one sub-repertory and another, the characters have weak hierarchy. This is illustrated in Fig. 17, where the repertory at the left is symbolically separated into three sub-repertories. The table at the bottom exhibits the strong hierarchy within any sub-repertory, and the weak hierarchy between the sub-repertories.

Actually, between the sub-repertories, Ranganathan's Colon Classification scheme behaves like a descriptor system. The cardinal product within any sub-repertory forms a tree, similar to that in Fig. 15, but when the cardinal product is taken of the whole repertory to get *P,* the tree character disappears. However, *P* is not a lattice, because the cardinal product partially ordered system *P* does not have any maximum element.

$$R: \left\{ \underbrace{\left[\begin{matrix} (abc) & \\ (ab) & (ad) & \cdots \\ (a) & (a) & \\ 0 & 0 & \end{matrix} \right]}_{\text{I}} \underbrace{\left[\begin{matrix} & (mpq) \\ (mn) & (mp) \\ (m) & (m) \\ 0 & 0 \end{matrix} \right]}_{\text{II}} \underbrace{\left[\begin{matrix} (xy) \\ (x) \\ 0 \end{matrix} \right]}_{\text{III}} \cdots \right\}$$

	Element	*Predecessor*	*Co-predecessor*
I	*a*	0	
	b	a	
	c	b	
	d	a	
II	*m*	0	
	n	m	Any of I
	p	m	
	q	p	
III	*x*	0	Any of I or II
	y	x	

FIGURE 17. The system of characters with hierarchy devised by Ranganathan effectively has the repertory divided into sections, and the characters have strong hierarchy within any sub-repertory, while they have weak hierarchy from one section to another. This is also shown in the table at the bottom.

In the preceding discussion, we have subjected various varieties of systems of characters with hierarchy to analysis. We have set forth the postulates for finding the repertory of partially ordered systems whose cardinal product becomes the space *P*. Although we have done this for a variety of systems, there are so many permutations and variations which can be taken on characters of this kind that there are surely more kinds of characters with hierarchy than have been described here. However, I believe that I have established the main outlines for the analysis of their language systems into the space *P*. This brings us to the discussion of the retrieval transformations from space *P* to space *L* of the document subsets.

When each document is entered into the collection, its subject content is analyzed and is delineated by a single point in the partially ordered system of characters *P*. Certain of the points in *L* represent subsets of documents which are all delineated by the same point of *P*. Thus, each point of *P* is represented by some largest subset of documents (including the null set). What is the same thing, each point of *P* is associated with a point of *L*. This association between the points of *P* and some of the points of *L* defines the transformation T_1.

One example of the retrieval transformation T_1 is in the method of classifying and storing patent specifications in the U.S. Patent Office. This method can be fully characterized as a system of characters with weak hierarchy employing the transformation T_1 for retrieval. It uses transformation T_1 because, for each point in *P*, there is actually a "pigeonhole" containing the specifications for that particular character. (I am omitting consideration of the important use of cross-reference specifications which are filed in various other pigeonholes to assist retrieval.)

For any mechanical retrieval system, the transformation T_2 is to be preferred. It can be defined in much the same fashion as was done for descriptors. Referring to Fig. 18, for any character *x* we find the class *X* of all characters which are preceded by *x*. This class *X* is indicated by the shaded portion of *P*. We then find the largest subset of documents such that all the documents in the subset are associated with a point in the class *X*. This subset of documents is represented by the point $x^\star$ in the lattice of document subsets *L*. The point *x* of *P* is associated with point $x^\star$ of *L*, and this association defines the transformation T_2. We may again observe that the class of all image points such as $x^\star$ does not exhaust all of the points of *L*. As is the case with descriptors, the fraction of the points of *L* that actually are image points is negligibly small. Again we can call this class of image points $L^\star$.

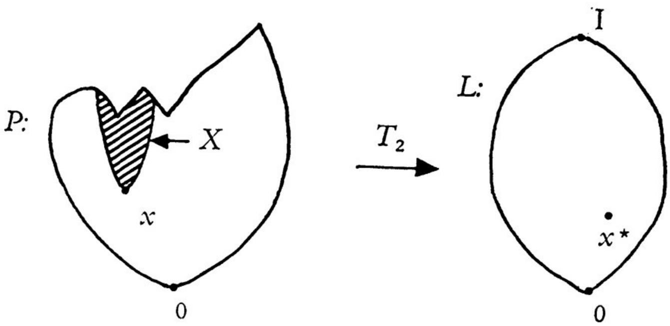
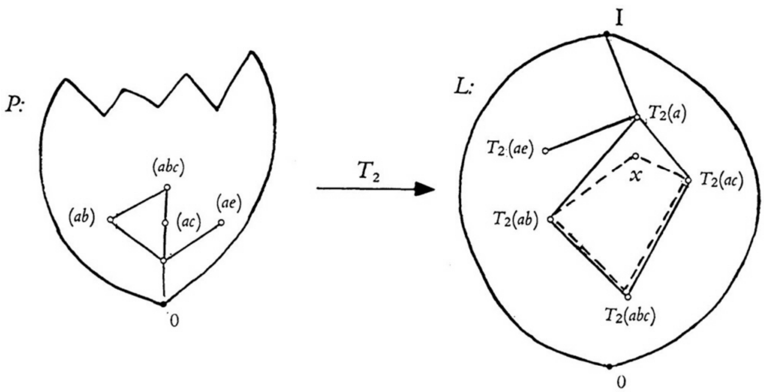


FIGURE 18. Illustration of the transformation T_2 for characters with hierarchy. The class of points *X* are all the points in *P* which are preceded by the prescribing point *x*, while the point $x^\star$ in *L* is the largest set of documents all of which are delineated by some point in *X*.

Consider now the behavior of transformations T_1 and T_2 with respect to an ascending line drawn in the partially ordered system *P*. Such an ascending line corresponds to starting out with the very broad or general retrieval prescription, and then refining it step by step until we have a very narrow prescription. As before, the image points of transformation T_1 are scattered points in *L*. They represent mutually exclusive subsets (except for the case in which the image points transform into the null element in *L*). Also as before, the transformation according to T_2 of an ascending line in P becomes a descending line in L. The line in L starts from the point I and goes down step by step until the characters in P become so narrow that there are no further documents, and then the image points are the 0 element of L.

Consider the transformations of cup and cap in P. If the system is one with strong hierarchy, the cup operation has no meaning in P because the structure of the partially ordered system P is that of a tree. Cap, however, is always defined. With weak hierarchy, cup is sometimes defined. When cup is defined, the cup and cap operations do transform by T_2 over into $L^\star$, although with an interchange of cup and cap in $L^\star$. As before, the cup and cap operations in $L^\star$ will generally give different elements than the cup and cap operations in L. A vivid illustration of this is shown in Fig. 19. With this observation, I shall end our discussion of characters with hierarchy.



$$\text{(I)}\quad T_2\,[(ab) \cup (ac)] = T_2\,(abc) = [T_2\,(ab)] \cap [T_2\,(ac)]$$

$$\text{(II)}\quad T_2\,[(ab) \cap (ac)] = T_2\,(a) \neq [T_2\,(ab) \cup T_2\,(ac)] = x$$

$$\text{(III)}\quad T_2\,(a) = [T_2\,(ab) \cup T_2\,(ac) \cup T_2\,(ae)]$$

FIGURE 19. Illustration showing that the cup operation in P is preserved by T_2, with cup going over into cap. However, the cap operation in P is not preserved by T_2 in L, though it is preserved in $L^\star$. The dotted lines show the cup and cap operations in the lattice L.

RETRIEVAL SYSTEMS BASED UPON CHARACTERS WITH LOGIC

By "characters with logic," I mean characters which are combined for the purposes of retrieval prescription by the so-called logical operations AND, OR, and NOT. These operations are respectively illustrated by the following possibilities: The AND operation upon the characters A and B, for a retrieval pre scription, requires that the selected document be delineated by both the characters *A* and *B*. The OR operation on the characters *A* and *B,* for retrieval prescription, requires that the selected document be delineated by either *A* or *B*. The operation NOT upon a character *A,* for retrieval prescription, requires that the selected document be delineated by the character which is the negative of character *A*.

The application of characters with logic to problems related to retrieval is very old, going back to 1884 in work on selection by Hollerith (10), and is closely related to the theories set forth in 1847 and 1854 by Boole (11), yet the matter still has some confusing aspects. Some of these have been discussed by Birkhoff (1, first edition, pp. 122–126). The application of characters with logic to problems of information retrieval has never been too well defined. There are many fuzzy edges. In this discussion, I can only point out what I think is a reasonable way to make use of characters with logic.

I wish to stress that one should not be overly impressed with retrieval systems based upon characters with logic merely because of the word “logic.” This word is used only because the characters of this method are combined in the same fashion as are the propositions in symbolic logic. We should remember that symbolic logic is a stylized view of things, and that the symbolism or method which is found useful in that discipline need not necessarily be the most appropriate symbolism or method for information retrieval. At all stages, we have to examine very carefully the retrieval interpretations placed upon any operations or symbols.

In order to bring systems of characters with logic into the framework of this discussion, we shall assume a partially ordered system, and we shall identify the operation AND with cup, the operation OR with cap and the operation NOT with the complement of an element. The operation of complementation needs defining. The complement of a character *A,* when it exists, is represented by $-A$, and is an element such that both $A\cup -A=\mathrm{I}$ and $A\cap -A=0$. These can be respectively interpreted as “the least upper bound of a character and its complement is equal to I,” and “the greatest lower bound is equal to 0.” The complement behaves somewhat like the negative or inverse of the element. We shall presume that, for every character, there is a complementary character.

We can combine the characters and their complements by cups and caps to get “polynomials” such as $[A\cup(B\cap C\cap -D)]\cup[A\cap D]$. The set of all such polynomials, which includes all the characters individually and their complements, and also the 0 and the I elements, forms a Boolean lattice. This lattice is the partially ordered system *P* in which each point (each polynomial) represents a possible retrieval prescription. Also, all possible prescriptions are represented somewhere in *P*.

An example of such a Boolean lattice, formed on the characters *A* and *B,* is shown in Fig. 20. All the possible retrieval prescribing polynomials that can be formed upon these two characters can be reduced to one of the forms shown in Fig. 20, and each is represented by a point in the diagram. Since this lattice is a Boolean lattice, its diagram is identical to, or isomorphic to, the lattice of the subsets of an aggregate. In the case shown in Fig. 20, it is identical to the lattice formed upon the subsets of an aggregate of four objects. Those objects correspond to the elements diagramed at the first level. Therefore, any point in the lattice can be attained by using the cup operation upon some set of the simple polynomials represented at the first level of Fig. 20. For this reason, such polynomials are called "minimal polynomials." (2, p. 322) Any possible Boolean polynomial can be formed by the cup operation upon one or more of such minimal polynomials.

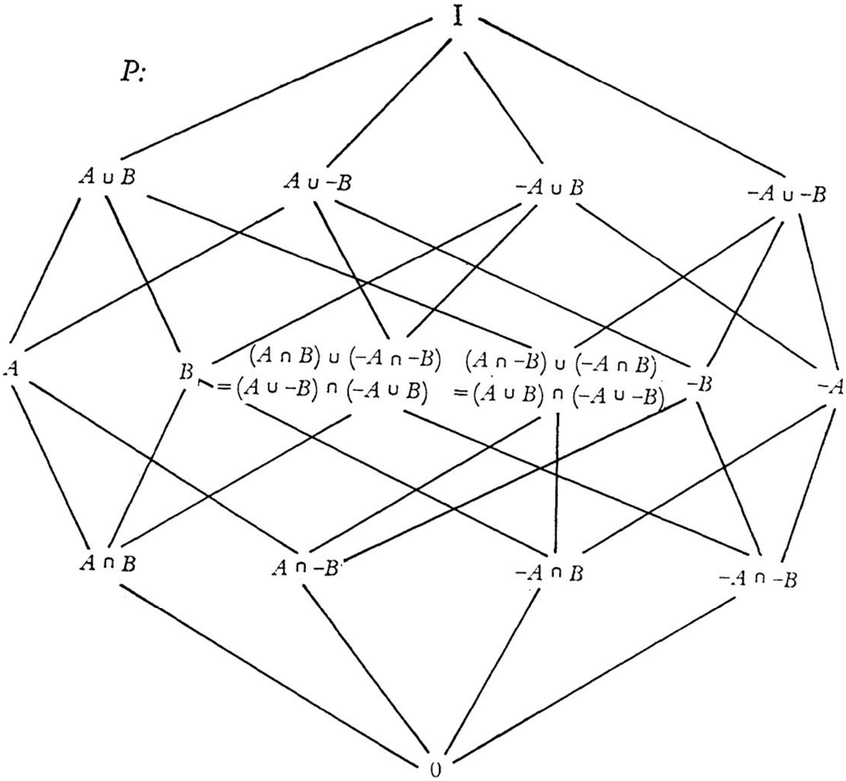


FIGURE 20. The Boolean lattice built upon the two characters *A* and *B,* by means of the operations cup, cap, and complement.

The lattice *P* can be generated as the cardinal product of simpler partially ordered systems. While there are a number of ways to factor *P,* the way that is the most interesting depends upon the use of the minimal polynomials. If there are N different characters in the system, a minimal polynomial is an expression like $A \cap B \cap -C \cdot \cdot \cdot \cap -G$ involving each of the N characters or its complement, all joined by a cap operation. No character is repeated or left out. There are 2^N such polynomials for N characters.

The repertory of simple partially ordered systems based upon minimal polynomials, whose cardinal product gives the lattice shown in Fig. 20, is shown in Fig. 21. The very close relationship between minimal polynomials and descriptors is obvious from this factorization.

Before going any further in the discussion of retrieval prescription, we must consider the problems involved in delineation by a single character. If the character A has the meaning “red,” and if a document describes something that is red, we use the character A to delineate the subject content of the document. If the object described is blue, then presumably we can use the character $-A$, meaning “not red.” If the document describes something for which the notion of color has no meaning, we are in a bit of a quandary. Presumably we should be able to sidestep the color delineation of the subject content. To do this, we would not use any of the color characters or their complements. If this is a valid viewpoint, a character can be asserted in delineation of the subject matter of a document only if the character is relevant to the document, and then only if it is determinable whether the character or its complement applies.

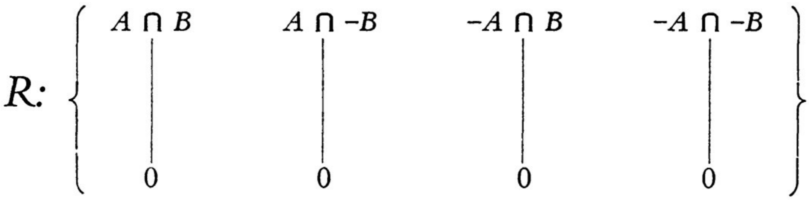


FIGURE 21. The repertory of minimal polynomials, whose cardinal product is the lattice shown in Fig. 20.

It is conceivable that a character may be relevant to the subject matter of a document although it is impossible to determine whether the character or its complement should be used in delineating the subject content. Then what do we do? Or, at the time of analyzing the document for delineation, it may not be known that a character is relevant although sometime later the character's relevance for this subject matter may be well established.

Problems of this general kind have never been properly faced by those who wish to employ characters with logic for retrieval. In what follows, I shall avoid these questions by presuming that at all times, it is evident whether or not a character is relevant, and if the character is relevant, then it is possible to tell whether the character or its complement should be used in delineating the document.

Consider now the operations AND and OR upon two typical characters *A* and *B* with respect to document delineation. The AND operation upon *A* and *B* is represented by $A \cup B$. This is a lattice polynomial. In delineation, it must be given the interpretation that both character *A* and character *B* are relevant to, and are descriptive of the subject matter of the document. (It should be noted that this is the way descriptors are used to delineate documents.) So far, so good.

However, the use of the OR operation in delineation is questionable. If two characters *A* and *B* are associated by the OR operation for delineation, they must be interpreted as saying that either one or the other character, or perhaps both, delineate the subject content. This is ridiculous. If a document is to be delineated, it should be delineated with precision, and not on the basis of such alternatives. For example, to delineate something by saying that it is "either red or square" (and this is easily possible with the use of the operation OR) is to be just a little precious, and not very helpful to a later customer of the retrieval system. For this reason, I do not believe that characters, or polynomials of characters, should be combined by the OR operation, or its equivalent, the cap, to form delineating polynomials.

According to this viewpoint, the only characters, or polynomials of characters, that may be used in delineation are: the null element 0, individual characters and their complements such as *A, B,* $-C$, and finally, polynomials of characters and their complements built up by use of the cup operation only. Thus, if the space *P,* as shown in Fig. 22, is the space of all possible polynomials, then (with the exception of the null element) all the points which may delineate lie either at the "equator" or above it. Reference to the simple case in Fig. 20 shows that not even all of the points at the equator are allowed in delineation. Not all the points above the equator are allowed either, since

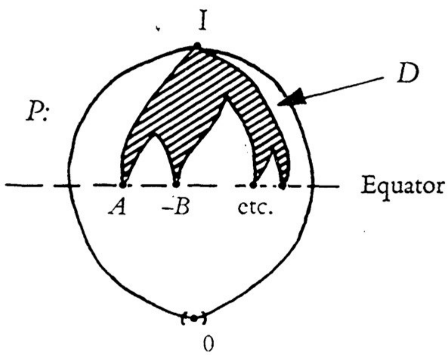


FIGURE 22. In a system of characters with logic, the points in the delineation space *D* (with the exception of 0) lie on the equator, or above the equator of *P*. *D* is a lattice, but it is not a sublattice of lattice *P,* even though the points of *D* are a subset of the points of *P*.

many of these points represent polynomials making use of the cap operation.

Let us call the set of all the allowed points, each representing an allowed polynomial, the delineation space *D*. The delineation space is a subset of *P* or is "embedded" in *P*. It is represented symbolically by the shaded area in Fig. 22. The *D* space of the lattice *P* in Fig. 20 is drawn out in full in Fig. 23. We should remember that the space *D* always includes the points 0 and I of *P*.

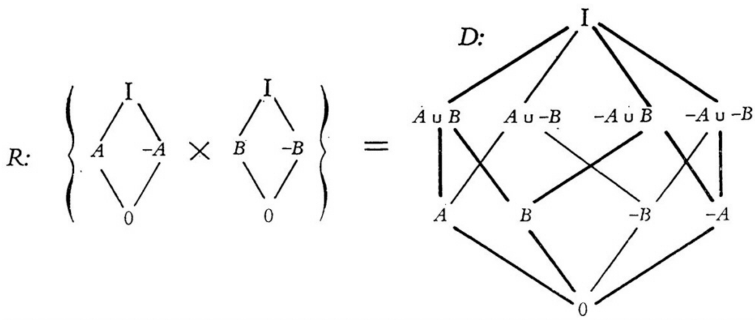


FIGURE 23. For characters with logic, the delineating space *D* is quite different from the prescribing space *P*. This diagram shows how the delineating lattice for two characters is generated. Note that lattice *D* is contained in lattice *P* of Fig. 20, but that *D* is not a sublattice.

Space *D* is a lattice, but it is not a sublattice of *P* because the cup and cap operations on the same elements give different results in *D* and *P*. Also, the lattice *D* does not have the same structure as the descriptor lattices, either for the case where there are as many descriptors as there are characters, or for the case where there are as many descriptors as there are characters and complements. Comparison of Fig. 23 with Fig. 7 will make this difference clear. The system *D* is not a Boolean lattice; it neither has unique complements, nor is distributive (1).

The lattice *D* can be factored into simpler partially ordered systems which, when combined by the cardinal product, give the lattice *D*. Such factors are shown at the left in Fig. 23 for the polynomial system of Fig. 20. Larger systems *D,* for a larger number of characters, can be generated by taking the cardinal product of a larger repertory of simple lattices of the same kind as shown in Fig. 23.

If the total number of characters is *N,* then the delineation of the content of the document may be according to any expression involving cup which has up to N different characters. Any such expression must never contain both a character and its complement.

If neither the character nor its complement is found in the polynomial of delineation, this means that either *(a)* the character is not relevant, or *(b)* it is impossible to determine whether the character or its complement is the appropriate delineation.

Although the content of a document can be delineated only by a point in the space *D* which is a subset of *P,* the retrieval prescription can be framed in terms of any of the points of *P*. This means that we can build up any combination of requirements involving OR, AND, and NOT. For retrieval, we can meaningfully ask for things that are either "red OR square." As another possibility, we can ask for those things that are "red AND square, or, if not that, are NOT long." This can be signified by $(A \cup B) \cap -C$.

This leads us directly into the discussion of the retrieval transformations for characters with hierarchy. The retrieval transformation T_1 is not very interesting; this is the transformation which associates each point *x* of *D* with a point $x^\star$ of *L* representing the largest subset of documents such that all the documents of the subset are delineated by the same point in *D*. Note that by T_1, the only permitted retrieval prescriptions are points in *D*. The image points in *L* will be much the same as in the previous examples. In particular, the non-null image points in *L* will represent mutually exclusive document subsets. The rest of the features of the transformation T_1 are also the same as they were for the other retrieval systems.

The retrieval transformation T_2 is more interesting and useful. Any point *x* in *P* may be a retrieval prescription. For any prescribing point *x,* we define the class *X* of points in *D* each of which is preceded by the point *x.* Then we look to *L,* and find the largest subset such that every document associated with the subset is delineated by one of the points in *X.* Let this subset be represented by $x^\star$. Thus, for every point *x* in *P,* there is an associated point $x^\star$ in *L,* and this association defines the retrieval transformation T_2.

A retrieval system based upon delineation according to a point in *D,* prescription according to a point in *P,* and the retrieval transformation T_2, can be a very useful system. However, the problems of relevance and indeterminancy of the characters still requires further study. Also the interpretation of the complement holds a number of hazards. In particular, if character *A* or its complement has not been used to delineate a document, this document should not be selected by the prescription NOT *A* merely because *A* is absent. Yet there have been proposals for retrieval systems in which such selection was intended.

For some purposes, we might consider a retrieval system in which both delineation and prescription would be formulated in terms of descriptors based on the minimal polynomials. The scope of meaning for each descriptor would be defined to be compatible with the alternatives given by one of the minimal polynomials. For most practical purposes, a retrieval system designed on this basis would not be too helpful. It would, for example, allow delineation of objects by "red OR square." However, if such delineations were allowable in a retrieval system, minimal polynomials could be used as descriptors. If this were done, the "logical" operations would effectively disappear because the whole system would behave like a descriptor system. Such a system uses only set combination (cup) and inclusion (partial ordering); it does not use the notions of OR, AND, and NOT.

In concluding our discussion of characters with logic, let us consider the behavior of cup and cap under the transformation T_2. As with the previous retrieval systems, the image points of the points of *P,* under the transformation T_2, will be only a small fraction of the points in *L*. Designate by $L^\star$ the set of image points in *L*. This system $L^\star$ is a lattice, and the cup and cap operations in the system *P* go over into cup and cap operations on corresponding elements in $L^\star$. The relation between *P* and $L^\star$ is not an isomorphism, even though most of the structure is preserved, but is a homomorphism (1, p. vii; 2, pp. 155, 349).

Again the lattice $L^\star$ is not a sublattice of *L*. For example, the set $T_2\ [(A\cup B)\cap(A\cup -B)]=T_2\ [A\cup(B\cap -B)]=T_2\ [A]$ will contain elements delineated by only *A,* while neither of the sets $T_2\ [A\cup B]$ or $T_2\ [A\cup -B]$ will do so. Therefore the cup operation in lattice $L^\star$ does not conform to the cup operation on the same elements in *L*. This exhibits another peculiarity about systems of characters with logic. The behavior during reduction of the polynomials in *P* does not conform to what seem to be common-sense requirements. Let a retrieval prescription be: (red AND round) OR (red AND not-round). Implicit in this prescription is the requirement that geometric shape is relevant. However, after reduction in the lattice *P,* this prescription reduces to mere (red). According to the original retrieval prescription, documents on red geometric shapes would be produced, but not all the documents concerned with the color red. By the reduced polynomial in *P,* documents on red shapes, red spectral lines, red pigments, and others, would all be produced. Evidently, this is another aspect of systems of characters with logic which will require clarification.

CONCLUDING REMARKS

This paper has presented a mathematical model for the theory of language symbols in retrieval and has applied the model to three different families of retrieval systems. I believe that almost all the retrieval systems now used in practice lie within the compass of these three families and their variations. In making this statement, I have in mind the great variety of actual physical methods and mechanisms which are used to instrument each of these families. Although this model, with its method of generating the inner details of retrieval systems, can represent most of the present retrieval systems, our theoretical task is not ended. So far, only the foundations have been laid. I am well aware that there still remain some chinks, yet to be filled, in these foundations. Nevertheless, these foundations are a basis for a program of further theoretical development, and I now want to discuss some of the topics that must be dealt with in the future.

The model as developed here does not incorporate certain statistical information about information retrieval systems that must be used in their design. For example, it is important to know the frequency with which each of the various characters are used. In our model, we can attach a "scalar" to each character in the repertory *R,* where the scalar is a number giving the relative incidence or frequency of usage of this character upon the documents in the collection. Also, in much the same way, each of the points in the system *P* is related by T_1 or T_2 to a certain number of documents in the collection. Thus, each point of *P* has two scalars associated with it, one according to T_1 and the other according to T_2. In the actual design of a retrieval system, we would like these scalar values on the points of *P* to be somewhat comparable at the same level in *P,* and to decrease in an orderly fashion as the level in *P* increases.

In the lattice *L,* there are also several important statistical features to be described. We have already mentioned one number of interest: the ratio of the number of points in $L^\star$ to the number of points in *L*. This ratio tells us the fraction of possible subsets that are actually used in information retrieval. Another statistic of *L* depends upon the fact that most of the lattice $L^\star$ is concentrated toward the bottom of the diagram of *L*. This means that most of the document subsets of actual use in retrieval are small. For system design, it is useful to know how the size of these subsets is distributed. A precise way of formulating this is to ask for the distribution of the number of points of $L^\star$ for each level of *L*. Another closely related kind of information has to do with the fact that sometimes a very simple prescription for information will produce the same number of documents as some other quite complicated prescription. This means that a simple prescription, represented by a point low in *P,* will transform into a point in *L* at the same level as some other point of *P* which is at a much higher level in *P*. Thus we are interested in a measure of the dispersion of the levels of *P* as a function of the level in *L*.

All the transformations of this mathematical model have presumed that the associated machines and codings are exact. However, certain coding schemes are not exact, and therefore we shall have to consider transformations such as T_3 which are based upon a probabilistic approach.

The elaboration of this mathematical model need not be restricted to mere embellishment of the diagrams with statistical information. Another very important virtue of this model is that it provides a method of generating the detailed structure of new language systems of retrieval by the use of cardinal products on partially ordered systems. By changing the partially ordered systems that go into the repertory, we can come out with quite different retrieval systems. For example, in our discussion of characters with logic, we avoided consideration of indeterminancies due to relevance, or to the applicability of a character or its complement. However, in the practical use of characters, we must always make allowance for all kinds of lack of information, ignorance, and error. When we do this, my study seems to indicate that each of the simple partially ordered systems of the repertory shown at the left in Fig. 23 takes on the character of the system shown in Fig. 24. The cardinal product of a repertory consisting of such systems will result in a delineating system D which is much more complicated than the one shown at the right in Fig. 23. All complications of this kind need to be investigated. In a similar manner, repertories involving other simple diagrams can be set up. Such diagrams can express various useful properties of the characters employed.

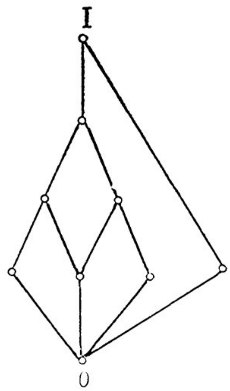


FIGURE 24. In a system of characters with logic, the relevance of a character may be in doubt, or it may be impossible to tell whether to apply the character or its complement in delineation. Also, errors may be made in delineation. In order to take into account these situations, the basic repertory lattice for a character and its complement apparently takes this form.

In regard to any program of generating new retrieval systems, we should remind ourselves that the discussion in this paper has dealt only with systems of characters which combine by formation of sets and classes. All such characters are commutative. With commutative characters, the order in which we make statements does not matter: *AB* means the same as *BA*. Commutative systems operate very well when we are dealing with objects in which no particular ordering or structure of the subject matter is involved, but this is not always true. The subject matter may be structured, either in a linear fashion, or in some form of a network.

If the information in a document is structured in linear sequence, the ideas are not commutative. It may then be useful to use characters in delineation and retrieval which themselves are not commutative. We may say that such characters combine by concatenation. As soon as we permit non-commutative characters, the partially ordered system of characters takes on an infinite number of levels, even though the number of different characters is finite.

Characters of still another kind are combined in a network fashion. We can say that these characters combine by reticulation, and such characters are used in chemical ciphering methods.

Another line of development involves the treatment of "super lattices." This technique seems to be called for when a document describes a number of different topics, and each topic is delineated by a separate point in the lattice *P*. The document as a whole should not be represented merely by the greatest lower bound or the least upper bound of these points. Instead, we should somehow be able to form a new kind of a representative point which keeps these various points of *P* distinct. The way to do this is to use the entire set of points of *P* as an aggregate of objects upon which a lattice is built by the formation of subsets from this aggregate. This defines the super lattice. Any document of however many different topics can then be fully represented in terms of a point in this super lattice.

Thanks are given to Mr. Russell A.Kirsch, of the National Bureau of Standards, for his specific encouragement in the development of this mathematical model of language systems of retrieval.

REFERENCES

1. BIRKHOFF, GARRETT, *Lattice Theory,* revised edition, New York, American Mathematical Society, 1948.
2. BIRKHOFF, GARRETT, and MACLANE, SAUNDERS, *A Survey of Modern Algebra,* New York, The Macmillan Company, 1941.
3. MOOERS, C.N., Zatocoding and Developments in Information Retrieval, *ASLIB Proc., 8,* 3–22 (February 1956).
4. TAUBE, MORTIMER, and ASSOCIATES, *Studies in Coordinate Indexing,* Washington, D.C., Documentation Inc., 1953.
5. FAIRTHORNE, R.A., The Mathematics of Classification, *Proc. British Society of International Bibliography, 9* [4], 1947.
6. FAIRTHORNE, R.A., The Patterns of Retrieval, *American Documentation, 7,* 65–75 (April 1956).
7. MOOERS, C.N., Zatocoding Applied to Mechanical Organization of Knowledge, The XVIIth Conference, Federation Internationale de Documentation, Rome, 1951; *American Documentation, 2,* 20–32 (January 1951).
8. MOOERS, C.N., Information Retrieval on Structured Content, paper published in the proceedings, *Information Theory—Third London Symposium,* edited by Colin Cherry, pp. 121–134, London, Butterworth's Scientific Publications, 1956. The full paper as presented at the meeting is published as *Zator Technical Bulletin* No. 102, Cambridge, Mass., Zator Company, 1955.
9. RANGANATHAN, S.R., "Colon Classification and Its Approach to Documentation," pp. 94–105 of *Bibliographic Classification,* edited by J.H. Shera and M.E. Egan, Chicago, University of Chicago Press, 1951.
10. HOLLERITH, HERMAN, "Art of Compiling Statistics" U. S. Patent No. 395,782 dated January 8, 1889, original application filed September 23, 1884; also "Apparatus for Compiling Statistics," U.S. Patent No. 395,783, same dates.
11. BOOLE, G., *The Mathematical Analysis of Logic,* Cambridge, 1847; *An Investigation into the Laws of Thought,* London, 1854 (New York, Dover Publications).

Abstract Theory of Retrieval Coding

CLIFFORD J.MALONEY

1. INTRODUCTION

The current feverish tempo of research activity in document retrieval has been accompanied, if not caused, by three other developments, (1) a rapid increase in the record to be controlled, involving new terminology, taking new forms, and appearing under widely dispersed auspices; (2) an increasing need to employ finer detail, to incorporate the new terminology more promptly, and to retrieve documents from unanticipated points of view; and (3) technical developments in the area of information theory, of information processing equipment, and in library science and its tools.

Proposals for more efficient and more powerful literature retrieval have taken a number of different forms. The present paper involves an attempt to develop a general theory adequate to afford a basis for an objective comparison of alternatives. While its main focus is on superposition coding, attention must be directed to virtually all phases of the problem because of their bearing on the superposition aspect.

The paper is concerned with developments involving the application of information theory considerations to the design of mechanical document retrieval systems, particularly those embodying superposition coding. A preliminary discussion of this problem was presented by title at the 1954 Annual Meeting of the American Statistical Association in Montreal, Canada. A description of a specific machine retrieval system whose development motivated the present study has appeared in *American Documentation* (1).

2. OUTLINE OF PROBLEM

The overall documentation problem involves: (1) the acquisition and preservation of documentary items, (2) the reception of requests for items involving subject matter and incomplete bibliographical description (or none) and the

CLIFFORD J.MALONEY Fort Detrick, Maryland.

identification of the item from this description (largely the subject matter) and (3) fulfilling the requests, by supplying copies on loan, microfilms, bibliographies, abstracts, or in other ways. The present paper is concerned only with the second stage. Before the advent of mechanical aids (including edge-notched cards), subject control of documents proceeded largely by the proliferation of an ordered set of index entries, primarily in card form. It was recognized that a considerable element of "scanning" or "searching" remained, however. The early machine developments, particularly the edge-notched cards, led to greater stress on scanning and less (or none) on preordering of the file, though this latter has always been an essential ingredient in the C.B.C.C. system.

Unfortunately, no scanning system so far developed has had such a high scanning rate at such a low cost that complete dependence could be put on this approach, though one has been projected (4). Current trends of technical developments seem to emphasize the perfection of "random access" storage facilities, which, then, may reestablish in the new machine technology some, at least, of the advantages of ordering afforded in the traditional card catalogue.

The long scanning period of currently available equipment and its cost, on the one hand, and the preparation and storage costs of complete card indexes on the other, have led to a study of superposition coding by Mooers (5), Wise (6), and Isbell (7). These authors use somewhat different approaches, and each differs from the one under study locally (1). A general theory seems appropriate to compare these and other approaches.

3. INFORMATION THEORY ANALOGUE

The analogy of this problem and that encountered in communication has been frequently noted (1–3), but it is believed that attention has usually been directed to the wrong point in the communication analogy. In the latter a "source" produces a "signal" which is coded, transmitted, decoded, and received as shown in Fig. 1. It is believed fruitful to consider that the problem which superposition coding attempts to solve is that of coding a signal source whose

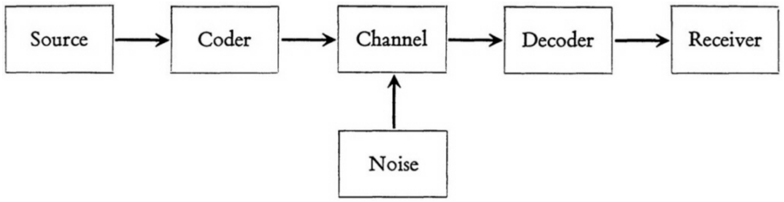


FIGURE 1. Schematic diagram of idealized communication channel.

rate of signal generation exceeds the channel capacity, and that the resulting equivocation is due to the form of signal compression by the encoding process adopted, not to channel noise. Its communication theory analogue would then, more accurately, be cross talk. Noise interference exists, of course, produced by human and machine errors in the operation of the system.

One consequence of viewing superposition coding as a signal encoding problem is that it raises questions of (1) whether other previously unthought of devices are also possible, and (2) the relative efficiencies of various forms of superposition coding. It will be shown that superposition coding *is* the only possible signal compression system involving the use of two-state devices, though it can be generalized when multistate devices are employed. By contrast, noise must be taken as given.

4. BUILDING BLOCKS

Subject retrieval of bibliographical items involves an informational content expressed in a written or at least recorded form of a specific language, or system of symbols. A purely formal treatment would start with this material (including punctuation, word order, alphabets, and type styles) and derive all details in successive steps. When applied to spoken language, this procedure has been worked out rigorously (9). Some work has been done on the written form of the language (10), though relatively little. Yngve (11) has done work paralleling the work of structural linguists, apparently independently.

What this work shows is that the “meaning” of any sign or sound complex is given by the set of all contexts in which the complex is found. The length of context to be considered in any case can be taken as the minimum length common to repeated occurrences. Different sign or sound complexes which share all contexts are identified, in the case of sounds into phonemes and then morphemes, in the case of written text into letters or diphthongs and then syllables and words. The resulting words, idioms, and phrases, together with punctuation, type style, and word order, express the entire information bearing content of the text.

The retrieval problem (step 2 of the first paragraph of this paper) consists of selecting from the total information-bearing content of each document a fraction sufficient to identify the utility of the document on the basis of later subject enquiries. In general, punctuation and type styles are ignored. Further, “little” words are in most contexts without interest. Some difference of opinion exists as to the value of ignoring grammatical form and word order. These points will be discussed more fully.

The result of this step is that the information content of each document is given by a selection of words, for which grammatical form may or may not be distinguished, and among which word order may or may not be differentiated.

5. ABSTRACT DEFINITION

A general theory must necessarily be able to treat all possible, or at least a large proportion of various coding systems. Further, the several systems should correspond to natural transformations of the theoretical model. Finally, few uninterpretable conclusions should result, though such examples are by no means unknown in applications of conventional mathematical disciplines. Explicit formulation of the document retrieval problem may proceed as follows. To each document in the file assign a "token" A, B, C, D, etc. No two tokens will be identical, since then the documents would be identical. A particular retrieval operation will consist of specifying or enumerating some subset of the totality of tokens. Call this operation a selection. A selection is a specification of the subject matter characteristics of wanted documents in terms of the properties of a query token. Now, in applying the selection to a particular record token, say H, it may or may not be possible to decide whether or not the selection criterion is fulfilled, i.e., whether or not the document is wanted, that the law of excluded middle applies in a somewhat extended (or distorted) sense. Critics affirm that this condition is widespread in conventional subject catalogues. Such a system will be considered "inoperable." The system may be made operable in either of two ways. The tokens and the selections may be so revised that a decision becomes possible in all cases. Or the rule may be adopted that all cases of doubt are to be selected (or rejected). Selecting all doubtful cases appears a reasonable solution in a document retrieval situation, provided they are not numerous, since the selected documents are to be subjected, at relatively little expense, to intelligent human scrutiny.

By either device it will be assumed that given any selection S_1 and any token G, it is possible to tell whether or not G meets the selection criterion S_1. Let S_1 in turn be applied to all the tokens in the system. Then the universe of tokens will be divided into two sets, the selections and the rejects. Let the class of selections be represented by A_1 (for admitted) and the class of rejections by R_1. Then every token is a member of either A_1 or R_1, but not both.

Next, let a second selection S_2, also with the excluded middle property, be applied to the collection of record tokens. S_2 forms two classes of tokens, A_2 of selections and R_2 of rejections. Now, it is possible that $S_1=S_2$ in the sense that $A_1=A_2$, i.e., they select the same concepts and hence the same documents containing those concepts and then $R_1=R_2$ and the two selections reject the same documents. A retrieval system in which every selection S_i selects exactly the same documents is of course trivial. Let then S_2 denote a selection in which $A_1 \neq A_2$. It may be true that $A_1 \supset A_2$ i.e., all tokens in A_2 are in A_1, or $A_2 \supset A_1$ or that A_1 and A_2 each include tokens not included by the other (whether or not they include tokens in common). To make progress we will need two further assumptions. Assume that the tokens of the retrieval system include all the selections that will be (or may be) used; that is, that if a token appropriate to a document should appear to be precisely equivalent to a selection provided for by the system, the concept could be entered into the system and would have that selection for token. Then S_2 is a member of either A_1 or R_1 and S_1 is a member of either A_2 or R_2. It is a matter of notation to assume that S_2 is in A_1. Of course it is in A_2. Moreover, all tokens in A_2 are wanted when the selection S_2 is used, and S_2 is wanted (is in A_1) when S_1 is used. It appears plausible that all of the members of A_2 and not merely S_2 are wanted when S_1 is the selection, and we will take it as our second assumption. This assumption is rendered more plausible on consideration of the meaning of the statement S_1 selects S_2. In a retrieval context this must mean that S_1 is a *broader* subject than S_2, so that S_2 is relevant to its treatment. Now, all documents selected by S_2 are special treatments of the topic of S_2, hence also of the topic represented by S_1. The selection relation thus has the three properties.

S_i selects S_i (1)

$S_i = S_j$ if, and only if, $S_i \supset S_j$ and $S_j \supset S_i$ (2)

If S_i selects S_j, and S_j selects S_k then S_i selects S_k (3)

Next, assume that every record token *J* appears in the set of selections; that is, every request is expressed in exactly the form in which some concept is or can be entered into the record. Then, the set of record tokens and the set of selections are isomorphic and each forms a partially ordered set (12). This result is normally accepted implicitly.

Two selections may be adjoined to the set, if not already included. The null selection selects no documents, and the universal selection selects all. With these two members added (if necessary) every "operable" retrieval system forms a lattice (12).

6. Mathematical model

In order to further study the abstract retrieval system defined in Sec. 5 an analytical representation is appropriate. On the basis of the structural analysis mentioned in Sec. 4, assume that all concepts in the system can be arranged in several sets or vectors

$$(X_1, X_2, \cdots X_m)$$
$$(Y_1, Y_2, \cdots Y_m)$$
$$(Z_1, Z_2, \cdots Z_m)$$

etc., where each member in any one vector can be substituted for each of the other elements in at least some contexts and give a meaningful expression (though presumably with an entirely different meaning). By assigning each "meaning" to a particular position in one of these vectors, any given concept can be expressed by assigning unit value to its associated position and zero everywhere else. The concept X_i becomes

$$(0, 0, 0 \cdots 0, 1, 0, \cdots 0)$$

A vector with a single unit and all other components zero will be called elementary. The zeros and ones are of course mere arbitrary signs at this stage, and in particular are not to be endowed with any numerical properties. Dashes could have been used for absence of concepts and a check mark for presence. Or the ones can be thought of as tally marks. The numerical quantities are used in the text because they will shortly be endowed with their usual properties. Likewise the expression "vector" is a mere name and endows these objects with no vector properties.

If one concept is followed in some context by a second concept from the same vector, or from another vector, this *ordered pair* of concepts can be symbolized by the matrix obtained as the axial product of the two vectors. The term axial product is being used to indicate that operation on two vectors, (or higher dimensional arrays) where elements are obtained as all possible products of pairs of quantities, one taken from the first vector and the other from the second. The arrangement in a two-dimensional array is merely mnemonic. Having regard to the non-quantitative role of the vector elements at this stage, the axial product of vectors can be regarded as the two dimensional tabular display formed by writing one vector across the top and the other vector along the left margin, and placing a unit (or check mark) in that cell which has a one *both* at its head and in its left margin, merely as an indication of joint occurrence. There will clearly be only one such instance. The axial product will appear as the matrix

$$\begin{Vmatrix} 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \end{Vmatrix}$$

Matrices of this form will be called elementary matrices. The *ordered pair* of these same two concepts in the reverse order would be symbolized by the transpose of the above matrix.

An ordered *triplet* would be symbolized by a three-dimensional matrix or tensor, obtained as the axial product of a vector and a matrix (or three vectors). In the same way *m*th ordered relations would be represented by *m*-dimensional tensors, containing a single unit in one cell and zeros everywhere else. The contents of any item in a document collection may be represented by a set of such tensors. If all the relations referring to a document are not of the same dimension or rank, each can be built up to that of the highest dimensional one by inserting zeros on added axes.

On the other hand, the argument of this paper can be followed adequately by thinking of two dimensional arrays, or matrices, only. Further, the matrix or tensor form is introduced for its suggestive power and to facilitate calculation by the rules for tensors. The arrayed collection of zeros and ones will normally be arranged in a linear or two-dimensional pattern for actual entry into the physical retrieval system.

In any retrieval system each of these tensors will be encountered with a certain frequency p_t where

$$\sum_t p_t = 1 \tag{4}$$

Each tensor can be thought of as a "message" in the retrieval system.

The number of possible tensors is enormous. Most will never occur. Shannon's channel capacity theorem says that it is possible to replace each by a "code" so that a maximum "transmission rate" is achieved. In the retrieval context the notion of "minimum message length" is more natural. To achieve minimum message length, however, in effect, requires a "memory" adequate to hold all possible tensors and their code "equivalents," say a serial number in the order of frequency of occurrence. This memory demand is impossibly great (as is in fact true in the communications case). Actually, it would be necessary to provide in the "code dictionary" for every permutation of the axes, since a prescribed order will not be available. The physical form of the "memory" in traditional library systems is the card catalogue. In mechanized systems it may be a file of punched cards, reels of magnetic tape, Minicards, etc.

The retrieval system design problem then is to so balance the three factors (1) "memory" or file size, (2) "message length" or searching time, and (3) equivocation or unwanted items that an optimum balance is obtained.

7. TENSOR COMBINATIONS

It has been indicated how relations of any complexity can be represented as *m*-dimensional tensors, with a single unit in one cell and zeros elsewhere. The linearized form of such a tensor would be extremely long, and hence the consequences of reducing the size is of interest.

A first simplification is possible. The representation of a document in the files is constructed in terms of a sequence of *m*th order tensors. It would be entirely possible to place all tallies for a given document in a single tensor, which would then have units or tallies in several cells, and zeros in all other tensor positions. Now, it is clear that this consolidation can be reversed and the several component tensors be written separately, given only the combined result, except that we will not know the order in which they originally appeared. If then, the dimensions of the tensors are sufficiently great so that no residual contribution from context is relevant to the retrieval problem, we can consider that the contents of a document are given indifferently by an unordered collection of elementary tensors (i.e., tensors with a unit in a single cell and zeros elsewhere) or by a single tensor with units in several cells and zeros elsewhere.

The analogy of the process by which mth order relational tensors were formed with axial multiplication has been previously noted. The analogy with tensor addition of the aggregation of the several tensors relating to one document with a single tensor of the same rank, but no longer elementary, is immediate, and suggests assigning the zeros and ones their normal numerical roles. In this connection it may be noted that just as in tensor addition, the element in the *ijk*th cell of the sum depends only on the elements in the corresponding cells of the summands.

Before completing the identification of our zeros and ones with those of arithmetic, the selection process should be examined from this point of view. In tensor language we will be given a “record tensor,” i.e., one from the files, and a “selection tensor,” i. e., one expressing the concept sought, and we will wish to decide whether or not the record tensor meets the selection criterion. If both the record and the selection tensors are elementary, we can select by identity. If, however, all elementary tensors relating to a document have been consolidated into one by tensor addition, the equality criterion will cause rejection. If attention is directed to the one cell of the elementary selection tensor containing a tally, it is clear that the document is wanted if the corresponding cell of the record tensor contains one or more tallies, otherwise not. In logical terms this means that, if the “class” of tallies in that cell of the record tensor “includes” that in the same cell of the selection tensor, selection is desired. In numerical terms this desired result is effected by the ≥relation. If we indicate the number of tallies in this cell of the record tensor by r and the corresponding number for the selection tensor by s, then the selection criterion is

$$r \geq s \quad (5)$$

In this case, r and s are necessarily unity, but a general symbol is desired for the argument to follow.

Since the selection tensor has zeros in all its other cells, relation (2) is necessarily satisfied by them all. Hence, we can apply the criterion to the record and selection tensors as wholes.

At this stage it is presumed that any query that may be phrased is represented by an elementary tensor in the system, hence that, in all cases, the selection rule can be

$$R \geq S \tag{6}$$

where the elements of the tensors R and S are positive integers or zero, and S is elementary.

It should be noted that, despite the fact that the inclusion rule and not the equality rule has been used for selection no false selections or "drop-outs" are obtained. This is due to the fact that each unit or tally in the record tensor unambiguously signifies a complete concept. If the tally is present in that cell, the concept is discussed in the document concerned irrespective of its other contents. There is then a reduction in file size (represented by the number of tensors in the system, which by this process has been reduced to the number of distinct documents) without causing false drop-outs. Tokens for the individual record tensors would, however, be extremely long, consist almost entirely of zeros, and have a high level of redundancy.

There is another price. When each document was represented by a full set of distinct elemental tensors, selection could be by the equality relation. Hence, all elements in the selection tensor could be specified. Accordingly it was possible to order the record tensors. When the latter are combined into a single nonelementary tensor, such tensors being brought together by addition of each of their several unit cell entries, a uniquely determinable order is not possible and selection must be by searching.

Generic searches are possible in either situation if elementary tensors designating genera are introduced into the system, or if searches are made on every species in the sought genus.

8. UNIQUENESS OF THE SELECTION PROCESS

In Sec. 7 we established in terms of a tensor model of a document retrieval system that any "operable" system must constitute a lattice, the same result arrived at in more general terms in Sec. 5. Accordingly, any conclusions arrived at by study of the tensor model employing its lattice properties only, necessarily hold for any possible retrieval systems. In the tensor model, we actually are presented with two isomorphic lattices, one of the record tensors and one of the query tensors. The two are related by a "selection" operation, the ordinary ≥of arithmetic. As this relation is uniquely applicable to a lattice, any other selection relation would have to be isomorphic to it.

We have further seen that it is possible to combine separate record tensors relating to a single document, and yet retain unique selection under the ≥relation. In the tensor model this combining operation takes the form of tensor addition. In order to permit combination, or addition, of tensors of different dimensions or rank, the operation of bordering the smaller tensor with zeros is introduced. Such tensors will be called "augmented." It is clear that it is still possible to border tensors with zeros even though the tensors to be combined are of the same dimensions and rank to begin with. If this is done we get what is analogous to, but not quite identical with, the "outer" sum or "direct" sum of the tensors.

The direct sum of *A* and *B* is

$$\begin{Vmatrix} A & 0 \\ 0 & B \end{Vmatrix}$$

while the direct sum of *B* and *A* is

$$\begin{Vmatrix} B & 0 \\ 0 & A \end{Vmatrix}$$

It will be seen that the direct sum is thus order preserving. From the sum we can recover, not merely the separate addends, but the order in which they entered the sum.

This device, however, has no attractiveness as a method for reducing the size of a retrieval system. For the order of the sum is the sum of those in the addends, and the number of elements increased, not decreased. Let *A* and *B* be bordered by zeros in such a way that they become:

$$\begin{Vmatrix} A & 0 \end{Vmatrix} \quad \text{and} \quad \begin{Vmatrix} 0 & B \end{Vmatrix}$$

then the ordinary matrix sum becomes

$$\begin{Vmatrix} A & B \end{Vmatrix}$$

This result (which we are calling the augmented sum) possesses all the properties of the direct sum of matrix theory under addition and subtraction (or selections). From one point of view it can be regarded as merely a different way of writing an ordered set of individual tensors, since the number of elements is unchanged, and all the information in the separate set is retained in the sum.

We have shown that record tensors in a retrieval system may be combined by addition and by "augmented" addition. We now show that no other combining rule is possible, or more properly that any such rule must be equivalent. For, let *A* and *B* be combined by any rule to produce *C*. Now *C* must be selected when either *A* or *B* is used as query. If *C* is of different order from *A* and *B* it must be of higher order, and cannot exceed the sum of their orders. For if of lower order the relation $R \geq S$ would not always hold even if *R* were bordered before the operation, and if of higher order, excess capacity would be present in the record tensors which would never function in selection. Hence, *C* must contain a unit in the same cell as *A* does, augmented if necessary, and the same for *B,* though the border zeros will be added to *B* in a fashion appropriate to this comparison. If *C* contained a unit in a cell corresponding to *neither* that for *A* or *B,* it would be selected by a query tensor, say *D,* to which the document did not relate. As this undesirable result buys no compensating advantage, such a combination rule for *A* and *B* would necessarily lower the efficiency of the system, and in that sense summation, ordinary or augumented, is the only combination rule applicable to the tensor representation of a retrieval system.

9. TENSOR REDUCTION

The preceding discussion concerned one way in which the "memory" or file size of a retrieval system might be reduced by reducing the number of tensors relating to any one document at the cost of introducing a new operation, file searching. This new operation, in the present state of technology involves equipment which is bulky, expensive, of low capacity, and/or time-consuming. Progress in reducing these limitations is proceeding rapidly. But, the present discussion, being theoretical, applies, so long as the searching process has not been reduced to the vanishing point.

The preceding section described how the file size or memory requirement of a retrieval system could be reduced to one tensor per document. Further reduction therefore can come only from reduction in the size of the individual tensors. Four devices are of common application in reducing the capacity requirements of individual tokens in a retrieval system, which in the present model means reducing the size of the individual tensors.

The first is merely to restrict the system to a special purpose. This has the effect of eliminating a number of items from the axial vectors altogether, and hence reducing their order. It reduces the retrieval problem involved to the same extent, thus making correspondingly less necessary the other tensor size reduction devices. The cost is that the special code resulting is not compatible with general codes or with other special codes, thus making accession separate operations in each system. The advantages of centralized cataloging and/or broad classification as tools in acquisition have not been clearly demonstrated in the present manual organization of documentation, but this conclusion does not prejudice their suitability in a machine system. The device of special classi fication is otherwise largely out of the hands of the system designer, being a function of the nature of the retrieval design problem, and not a tool to be used in its solution.

The second tensor size reduction process also results in axial vectors of smaller order and involves entering species under suitable genera eliminating from the system provision for indicating the dropped species. This device too reduces the order of the axial vectors.

A third common procedure is to truncate complex relations to binary, or triadic relations. This has the effect of restricting the dimension or rank of the record tensors to matrices or tensors of rank three (or whatever). A relation involving more terms would be reentered until all terms appeared.

The fourth simplification is to ignore the distinction between relations exhibiting the same concepts, but in different order. Here, instead of eliminating vector components, or axes, we instead must seek the analogue of the process in adding units to the tensors in the system. In fact, a unit will be given for each point which can be reached by employing the axes in all possible orders. A three-element relation would thus lead to 3!=6 tallies or entries in the record tensor.

All three of the other tensor simplification devices can be viewed from this point of view as a unifying principle. Dropping components in a tensor can be viewed as *entering* a zero for this term *in all record tensors*. The terms will of course then have *no* selective power which is the consequence of dropping them from the system. Entering species under genera is equivalent to entering a unit in the genus, *and zero in all its species* wherever any one of the species is encountered in classifying a document. Dropping axes is equivalent to entering zero *in every cell* of the dropped portion of the record tensors.

Conversely, we can disregard completely the model relation which these tensors bear to record tokens in an informational system and ask: "What natural rules would lead to accompanying entry of tallys in tensors by a set of symmetrically arranged satellite entries?" It is clear that one such rule would be to make entries in every tensor in some specified point, line, or plane wherever any tally is called for anywhere. This is a generalization of restricting the scope of the system. Another rule would be to enter zero in points, lines, or planes, but the identity of the last varying accordingly to the location of the tally entered in the given record tensor. This rule includes entering species under genera and limiting the rank of record tensors. Finally, each tally can be accompanied by a cluster of points *not* all in any point, line, or plane. If this cluster is a "regular" one, it correlates with disregarding order in the relation entered into the retrieval system.

In summary then, adding zeros in cells of a theoretically conceived relational tensor according to specified rules endows the set of tensors with the same informational properties as a retrieval system limited in an isomorphic fashion.

The consequences of these restrictions on the retrieval system in making selections in that system may be examined in purely mathematical terms on the modified record tensors. It is found that no essentially new device will be produced. This is a tribute to the ingenuity of preceding library and documentation investigators. If valid, the present essay establishes that procedures radically different in principle would appear to conflict with seemingly essential characteristics of an informational retrieval system.

In each case there is a loss of the ability to make sharp distinctions in retrieval. It is presumed that the distinctions lost are of lesser importance than those retained. This effect may be due to two entirely different characteristics of the process. On the one hand, it may be true that only one of the amalgamated record codes occurs with any frequency, or at least, when retrieval is conditioned by additional specifications. Thus "animal as a friend and companion" suggests a horse, dog, or cat when viewed as a friend of man, boy, or old maid. The numerical measure of this redundancy is given by the summation of the frequency of occurrence of the unwanted items.

$$\sum_{c \neq j} p \qquad (7)$$

where *j* is the one item in the combined group which is desired.

A second way in which the loss of discriminating power is not too great is that the amalgamated classes are all at least related and hence may be worth examining for possible relevance.

There will be no false drop-outs or selections owing to chance whether all truncated tensors are combined by augmented or ordinary summations.

10. COORDINATIZATION

Preceding sections have outlined a procedure employing a system of tensors as a model of any possible retrieval system. The number of bits required to hold each tensor is given by the product of the orders of the dimensions. Both order and dimension can be reduced by (1) restricting the field of applications of the system, (2) use of generic terms, (3) limiting the degree of relationships in the system. All tensors relating to one document can be combined into one by tensor addition without loss of discriminating power, but at the cost of requiring a searching operation in retrieval.

Returning to the consideration of the separate tensors in which order is distinguished, each will consist of Πd_i cells where d_i is the order on the *i*th dimension, each cell containing a zero except for one cell, which will contain a unit. It is of course universal to choose a different form of record, which we will regard as a coordinatized form of these tensors. Instead of a single tally in one cell of an *m*th dimensional vector, a tally will be entered in each coordinate vector of the tensor. The bit requirement is thus lowered from Πd_i to Σd_i, and now *m* tallies are used rather than one. Queries must be expressed in the same form, of course, and will consist of *m* vectors, each with a single unit. So long as the coordinate vectors relating to one concept for each document are kept separate, no random or specious drop-outs occur, and this still holds after augmented addition.

If order is not distinguished, each coordinate vector will have *m* tallies rather than one, and selection will produce analogous results to the extent that a query for any order will select records from the file for all possible orders. Whether this is tolerable or not depends on the given system. Mooers (5) has shown how to avoid this consequence if worthwhile. If the present system is a general one, Mooers' device must fit into it.

The linearized form of the full tensor representation which, it will be recalled, has a tally in a single cell, and zeros in all others can be viewed as a form of direct coding. The consequence, that neither augmented nor ordinary summation leads to false drop-outs in this circumstance shows that as long as concepts can be entered into a system as a single tally, by direct coding, false dropouts will not occur.

When however, the coordinate representations for each of the several concepts relating to one document are combined by ordinary addition, the phenomenon of false drop-outs arises. We have not strictly speaking defined ordinary summation of the coordinatized form of a tensor, but will do so as the ordinary addition of corresponding coordinates. It is immediate that this operation retains the properties needed in a retrieval system.

In a lattice the union of two elements *A* and *B* is not unique and may be obtained as the union of two other elements *C* and *D*. As the union has the ordinary inclusion property with respect to *any* element leading to it, the union of *A* and *B* includes either *C* or *D,* and vice versa. This, in the retrieval situation, is the false drop-out phenomenon.

A preliminary operation is available prior to coordinatization of the retrieval tensors. If two (or more) vectors are combined not by axial multiplication but by the conventional form of direct vector sum, the result is a vector of order equal to the sum of the orders of its factors, and which provides for dyadic relations (or the dyadic prefix of higher degree relations). This is the mathematical representation of what Mooers does.

Finally, what is really the converse of this process, the positions on any axis can be grouped arbitrarily according to some rule, and tallies placed in a posi tion within each group to designate each concept for that axis. Since, again, single concepts are represented by multiple entries, ordinary addition will lead to false drop-outs on the criterion.

Multiple tallies may be assigned to single (if complex) concepts in either of two ways. If done arbitrarily, attention can be given to minimizing coincidences upon ordinary summation, which means minimizing false drop-outs. If done according to some rule, one rule can be to convert the lattice into a serially ordered set (which we know can always be done). That conversion may be accomplished by classing together all elements for which the lattice relation holds, then arbitrarily choosing a new element, not in the chain, and forming a new chain, and continuing until all elements are in some chain. Then the chain can be numbered serially, and the number of any element will be given by the number of its chain, and its position number within the chain. If now chain numbers are direct coded, then they will not overlap on ordinary summation and hence false drop-outs will *not* occur as to chain number on selection by ≥. This will be a classified code.

11. DERIVATION OF DROP-OUT PERCENTAGE

Treatments of the problem of the number of drop-outs which will occur under superposition coding are given by Wise (14) and by Mooers (5). Following Wise, let

X=number of concepts entered in the record

H=number of positions in coding subfields

G=number of positions with an entry

Y=number of subfields

F_d="dropping fraction" or number of selected record

M=number of concepts specified by a query

Wise gives the formula for the number of drop-outs, assuming (1) random coding of concepts, and (2) random association of concepts in documents as

$$F_d = \left(\frac{G}{H}\right)^{YM} \quad (8)$$

A treatment by the use of the method of this paper would proceed by seeking the interpretation in terms of elementary tensors of a record token. If G_i positions are occupied in each of the Y coding subfields of the record token then all X concepts can be reconstructed by selecting a position from each subfield in turn. Additional concepts, however, will also be given, by this process, since the number of all possible selections, one from each subfield, will in general exceed *X*. The number will be

$$\prod_{Y} G_i \tag{9}$$

The number of possible codes in the system is given by H^Y. Each record token divides this total possible number into two portions, one, given by (9), of concepts which will select this token, and one containing all that will reject it. Let a selection be specified. Then its *A* class will contain all records for which its code is included in the set (9). The probability of this occurrence in complete generality is well known, see Uspensky (15), but will not be used here.

A second selection will divide the entire record collection in an analogous fashion. The two selected classes A_1 and A_2 will include records in common if documents in the file treat both subjects, or if both codes can be formed from entries in one document. Both cases are covered by the requirement that the code combination for either selection code be contained within the selected class of the other, all those combinations not present in both selection classes being rejected.

The probabilities associated with these events are given by those for occurrence and of joint occurrence of concepts in the material in the file, and hence are in principle computable on the basis of "language statistics."

Comparison with Wise's special case is obtained by taking all probabilities as equal and independent and using expected values. In this case the residue class is given by G^Y for a single selection. All possible classes are given by H^Y, and hence the ratio of the number of items in the selection class *A*, to the total number of possible codes in $G^Y/H^Y = (G/H)^Y$. In the case of independent selection by M codes this leads to $(G/H)^{YM}$.

CONCLUSIONS

A. Any retrieval system forms a lattice. More precisely the set of record tokens and the set of query tokens, real and potential, form isomorphic lattices.

B. The tokens of either set can be represented as *m*-dimensional tensors. If separate tensors are devoted to each *m*-dimensional (or lower) relation in each document, then every tensor in the system will be elementary.

C. Only two methods are available for combining the tensors for the several concepts relating to a single document, ordinary and augmented tensor addition.

D. If the full *m*-dimensional tensors are combined in either of the two admissible ways, by ordinary or augmented addition, selection will be unique.

E. If combined by augmented addition, searching will be required, and a very large memory capacity, file size, as well. Selection may be by equality or partial ordering.

F. For tensors combined by ordinary summation, only one possible selection relation is available for selecting documents dealing with a subject represented by a query tensor, the lattice partial ordering relation ≥; memory requirements are reduced to one tensor per document but ordering is not possible, so that searching is required.

G. Tensor size may be reduced by (1) restriction of system scope, (2) use of generic rather than specific concepts, (3) disregarding order in relations, and (4) restricting the complexity of relations entered in the system. Each of these devices has a natural interpretation in terms of the tensor model.

H. Great economy in bit requirement is introduced by expressing record and query tensors in coordinatized form, with one tally per coordinate.

I. Coordinatized records combined by augmented addition result in no additional reduction in memory and require searching since ordering is not possible, but yield unique selections, either by equal or inclusion selection.

J. Coordinatized records combined by ordinary addition yield minimal searching length and file requirements but produce false drop-outs. The incidence of false drop-outs is measured by the combinations which can be formed from the coordinates initially entered into the sum.

K. Formulas for calculating drop-out fractions without assuming random coding can be based on "language statistics" appropriate to a given application.

REFERENCES

1. C.J.MALONEY. A Remington Rand retrieval system, *American Documentation* vol. IX (1958), p. 1–12.
2. KARL F.HEUMANN. Information theory in library and documentation activities, *LC Inf. Bull* vol. 13, Sept. 27, 1954, Appendix II.
3. R.M.FANO. Information theory and the retrieval of recorded information, in *Documentation in Action* edited by Shera, Kent, and Perry, 1956.
4. LOUIS N.RIDENOUR. Storage and retrieval of information, *Proc. Eastern Joint Computer Conf.* Nov. 7–9, 1955, Boston.
5. CALVIN N.MOOERS. Putting probability to work in coding punched cards, 112th Meeting, American Chemical Society, New York, 1947.
6. CARL S.WISE. Generalized punched-card codes for chemical bibliographies, 112th Meeting, American Chemical Society, New York, 1947.

7. A.F.ISBELL. A practical application of a punched-card system utilizing the superposition of codes, 114th Meeting, American Chemical Society, Portland, Oregon, 1948.

8. CLAUDE E.SHANNON. A mathematical theory of communication, *Bell System Technical Journal,* July 1948, p. 379; Oct. 1948, p. 623.

9. ZELLIG HARRIS. *Methods in Structural Linguistics,* 1951.

10. BOLINGERS. Visual Morphemes, *Language,* vol. 22 (1946) pp. 330–340.

11. VICTOR H.YNGVE. The translation of languages by machine, in *Information Theory,* 3rd London Symposium, Colin Cherry, editor, pp. 195–205.

12. GARRETT BIRKHOFF. What is a lattice? *Am. Math. Assoc. Monthly,* vol. 50, No. 8 (Oct. 1943), pp. 484–487.

13. F.I.MAUTNER. An extension of Klein's Erlanger program: Logic as invariant theory, *Am. J. Math.,* vol. 68 (1946), pp. 345–384.

14. C.S.WISE. Mathematical analysis of coding systems, Chapter 20 of *Punched Cards,* edited by Robert S.Casey and James W.Perry, 1951.

15. J.V.USPENSKY. *Elementary Number Theory,* McGraw-Hill, New York, 1954.

Maze Structure and Information Retrieval[1]

GERALD ESTRIN

MOORE'S ALGORITHM

In a paper (1) delivered by E.F.Moore, he established a group of algorithms for finding the shortest path through a maze. These algorithms resulted from pure research related to puzzles and games and an alertness to possible applications to complex communications systems.

The mazes considered were defined by a set of key points and connections between them. The links connecting them could be weighted by a cost function pertinent to the system represented by the maze. The simplest of the algorithms concerned a maze in which all links were of unit length.

A particular maze is illustrated in Fig. 1. The most elementary problem posed by Moore is: Given an origin point and a destination point in the maze, find a path from origin to destination containing the minimum number of links.

The algorithm established by Moore for this problem may be described in the following way:

Initial step. Select the origin point and tag it with a zero. Enable the recognition of the destination point whenever it may be selected.

Step i. Select all points which have not been tagged (i=1, 2, · · ·, k) but which are adjacent to the points which have been tagged and provide them with the tag i. Repeat this process until the destination point is selected and given the tag k. The tag represents the number of links in the minimum path.

Step j (j=k+1, k+2, · · ·, 2k). The second half of the process explicitly defines a minimum path starting back from the destination point. At each of these steps, select an adjacent point which has a smaller i tag. Provide that point with an additional tag, i'=i, and use the newly tagged point as the

GERALD ESTRIN Department of Engineering and Department of Mathematics Numerical Analysis Research, University of California, Los Angeles, California.

[1] The preparation of this paper was sponsored in part by the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

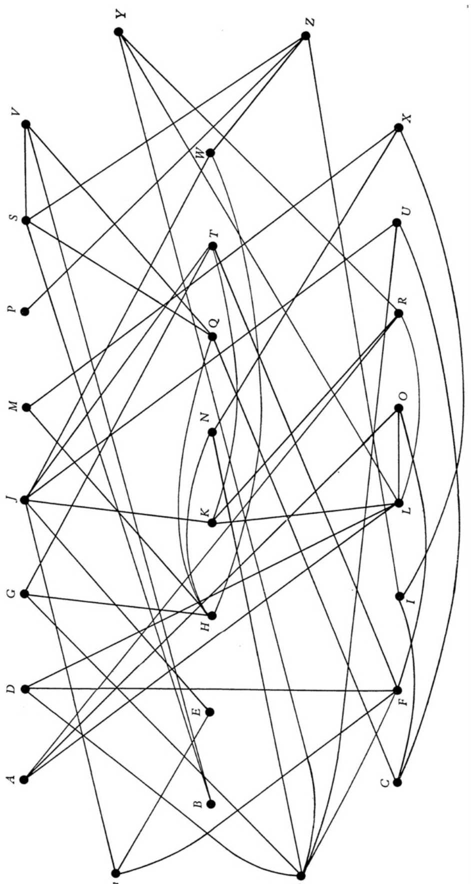


FIGURE 1.

base for the next step. Continue this process until the origin is reached and given the tag i'=0. All points carrying an i' tag lie on a minimum path.

In Fig. 2, the solution of the maze of Fig. 1 is displayed for the origin-destination pair *A–B*. There are two possible solutions containing seven links. These are *ALKJWZSB* and *ARKJWZSB*.

The simplicity of the algorithm makes it amenable to mechanization by existing digital techniques. One may, as usual, make compromises between the parallelism of mechanization and the speed or number of steps.

MAZE ORIENTATION

Moore's algorithm concentrates on the abstraction of a minimum path between two specified points in a maze. However, the first half of the process is unrelated to the destination point except for the last step.

If the network is regraphed in accordance with the step by step application of the algorithm, one achieves a reorientation of the maze with respect to the origin point. This process is carried out for the maze of Fig. 1 with A as the origin point and the result is displayed in Fig. 3. Comparison of Figs. 1 and 3 shows strikingly the relative ease with which the many paths may be followed after regraphing. For every point in Figure 3, the vertical distance from the origin represents the number of links between the point and the origin along a minimum path.

Figure 4 displays a graph of the same maze with respect to a different origin point and indicates how severely the structure of a network may change depending upon the point of entry.

INFORMATION RETRIEVAL SYSTEMS

Supported by the success of this very simple procedure for unraveling complex networks, examples of information retrieval systems are considered according to the following criterion: When the system is characterized graphically, can the maze organizing algorithm or a modification thereof be used to systematize the network with respect to search requirements?

THE SUBJECT CATALOG

The subject catalog achieves its maze attributes as a result of the sets of cross references linking the subject headings. Although it was developed as a search aid, the difficulty of keeping in mind much more than point to point search properties is a limitation to its usefulness.

Given an initial subject heading related to the search requirements, the

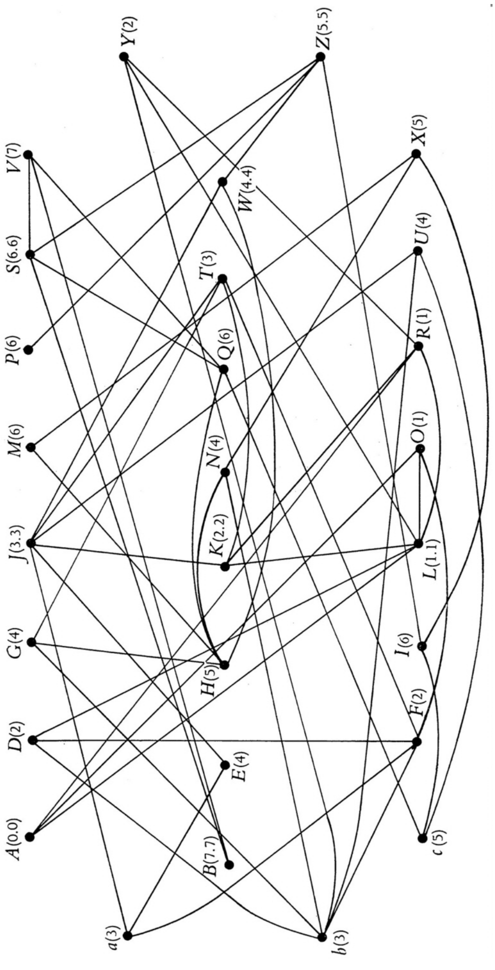

A(0.0)
D(2)
G(4)
J(3.3)
M(6)
P(6)
S(6.6)
V(7)
B(7.7)
E(4)
H(5)
K(2.2)
N(4)
Q(6)
T(3)
W(4.4)
Y(2)
Z(5.5)
a(3)
b(3)
c(5)
F(2)
I(6)
L(1.1)
O(1)
R(1)
U(4)
X(5)

FIGURE 2.

searcher would most certainly be helped in forming a strategy of search if he were able to study the character of the subject heading maze in a region around the point of entry.

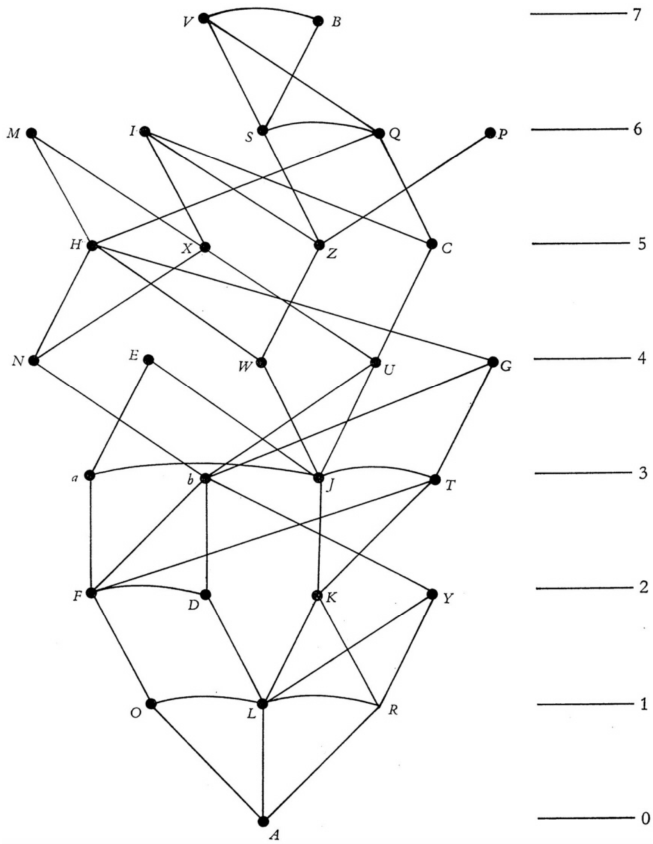


FIGURE 3. Maze of Figure 1 organized with respect to *A*.

If the point of entry is chosen as the origin point and the procedure described under "Maze orientation" is applied for a defined number of steps, it can be used to produce a graph of subject headings as linked by cross references.

A simple example is illustrated in Fig. 5. The searcher then chooses a most desirable path among the subject headings and proceeds through the subject catalog in a planned fashion. The graph of the localized region is extendable by selecting the extremity of a desired path as a new origin point.

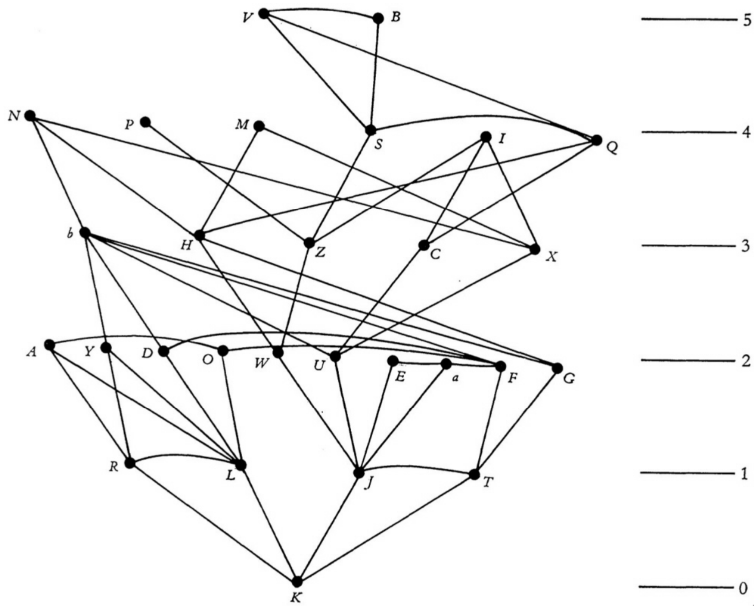


FIGURE 4. Maze of Figure 1 organized with respect to *K*.

Experiments leading to a preliminary evaluation of the effectiveness of this method are easily formulated. Two groups are assigned to a set of selected search problems. One of them follows conventional techniques, the other begins the search by graphing a section of the subject heading catalog as described. The number of references, pertinency, time of search, experience of searchers may then be included as parameters in determining whether to evaluate the engineering design of automatic devices to permit larger scale experiments. Such studies should be carried on in both special and general library collections.

MULTIPLE ASPECT SEARCH—THE PEEK-A-BOO SYSTEM (2, 3)

The Term-Document Index

The peek-a-boo retrieval system utilizes a card for each term in the index vocabulary and a fixed location on all cards for each document in the system.

It provides a highly parallel method for establishing the existence of documents which were indexed with at least the full set of terms defining the search request.

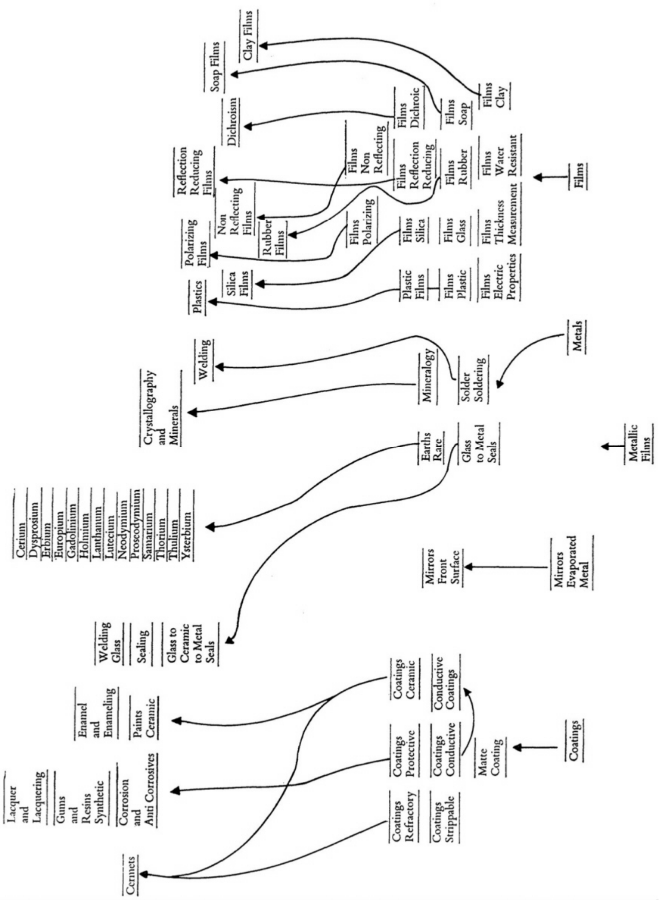


FIGURE 5. Graph of subject headings (4) derived from search objective "The properties of glass affecting the evaporation of thin metallic films on glass substrates." Initial subject heading: "Metallic Films."

One may visualize the system as having the highly linear and shallow structure of Fig. 6. The peek-a-boo technique is a superstructure which observes those documents which are intersected by at least all of the selected terms.

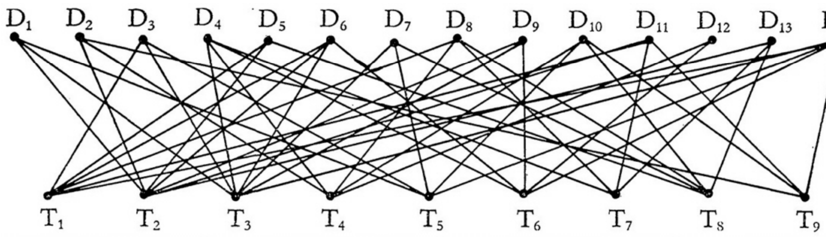


FIGURE 6.

The effectiveness of the search is dependent upon the choice of terms attached to each document when it enters the system and the choice of terms defining the search request. One may expect therefore the development of a maze of interconnections in the vocabulary itself to aid the description of ambiguities.

Three types of difficulty are obvious.

i. If no significance is attached to the ordering of terms, some of the documents may not be pertinent to search requirements. This type of difficulty is usually not too serious since the searcher can quickly reject such documents.
ii. If the set of search terms is not sufficiently specific, an excessively large group of documents may be retrieved. The search may then be continued with the addition of new significant term cards.
iii. If the set of search terms is overly specific relative to the collection, pertinent documents will be excluded—in the extreme no documents will be retrieved.

In the last two cases, it is evident that the effectiveness of this method will be dependent on one's ability to make perturbations of the initial selection of search terms. In general the way in which the selected set of documents converges to its final selection as term cards are added to the stack, is dependent on their order.

A point (set of documents) in the system is uniquely selected by the choice of the full set of search terms. However, if judgment is to be exercised about a perturbation on this set of search terms, it would be helpful to have some information about the nature of the system in the locality of the selected point.

The highly compressed graph of Fig. 6 may be expanded into a set of alternating term and document levels starting from a particular set of search terms.

One way of accomplishing this expansion is the simple application of the maze organizing method, using the complete set of search terms as a starting level 0; level 1 contains all documents linked to any one of the initial search terms; level 2 contains those vocabulary terms not already in level 0 but linked to the documents in level 1; level 3 contains those documents linked to the search terms of level 2 but not already contained in level 1; etc. This structure does not offer much help. Its first levels do not include the selectivity of the peek-a-boo technique, and therefore one cannot recognize a central region of the network corresponding to conditions resulting from the peek-a-boo operation.

Figure 7 displays an expansion of the network which has more selectivity in each of the steps and establishes with levels 0 and 1 the result of peek-a-boo selection. It is developed in the following way.

Level 0 contains the set of *n* search terms.

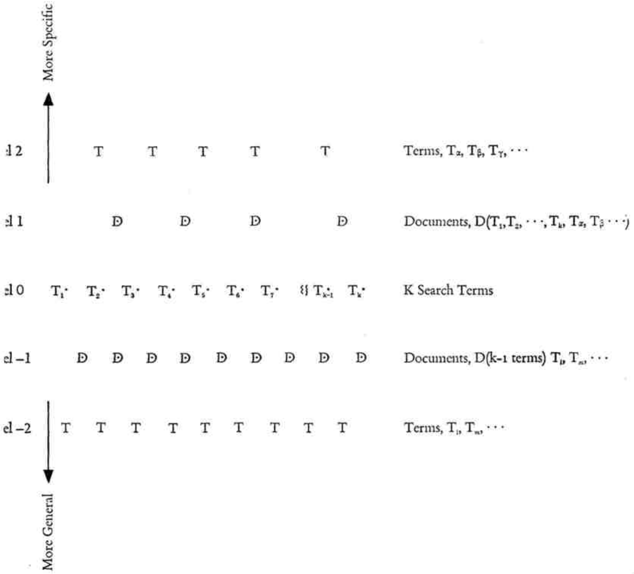


FIGURE 7.

Level 1 contains those documents indexed with *at least* all of the n search terms.

Level 2 contains the set of terms, assigned to the documents in level 1, but not already contained in level 0.

Level −1 contains the documents indexed with at least $n-1$ of the search terms, but not already contained in level 1.

Level −2 contains the set of terms assigned to the documents of level −1 but not already in level 0.

A further modification which might be made to increase the probable pertinency of documents in level −1 would require that k of the $n-1$ search terms must be included in the index set of all documents represented.

Levels 0 and 1 are obtained with the standard peek-a-boo technique. Level 2 implies the extraction of the index set for each document selected and the subtraction of the set of search terms. Study of level 2 will guide the choice of additional terms. Level −1 is obtained by the application of the peek-a-boo technique n (or $n-k$) times with a single term card withdrawn from the stack. Level −2 is developed at each of the steps of forming level −1 by extracting the index set for each document selected and subtracting the set of search terms. Study of level −2 will guide the choice of document selected in level −1.

As in the case of the discussion of the subject catalog, a set of small scale experiments may be defined to evaluate this procedure. If extensive cross referencing is utilized with the dictionary of key terms, the procedure outlined for the subject catalog may be used to aid in the selection of search terms.

The Term-Associative Index

In an attempt to enlarge the scope of the search and particularly to have it directed along "association trails" formed within the vocabulary, Taube conceived of a modification of the peek-a-boo system which would define a network of two-term associations such that, "the mechanical display of such networks of association effectively solves the challenge of Vannevar Bush and provides the 'coincidences' of ideas which Benier called the most important characteristic of an information system" (3).

In addition to the Term-Document cards discussed in the previous section, Taube established a card for each term on which there is a position defined for each term in the vocabulary.

N of these term association cards are stacked together (with the restriction that each term card added to the stack corresponds to a term which permitted transmission through the stack in the previous step). The peek-a-boo technique is applied and the outputs imply the existence of documents classed within the following extremes: single documents indexed by at least all the n+1 terms used or groups of (n+1) documents each indexed by a different one of the search terms.

Taube conceives of the use of this device in a man machine process in which the human selects the next term card from the outputs of the previous group. It is apparently Taube's hope that the “association machine” will lead to an optimum selection of search terms because the searcher is restricted to selection of terms from a group which has some possibility of producing a document. However, there is nothing in the process which either guarantees a high probability of retrieval or eliminates the problems raised in the discussion of the Term-Document Index. Taube's procedure may lead to a better selection of search terms than the less guided experience of the searcher but is unlikely to dispense with the need for a directed perturbation of that set of terms.

CONCLUSION

The above discussion has considered the structure of three retrieval systems in an attempt to define search aids which are capable of displaying the local properties of a retrieval system maze. The approach taken here has assumed existing retrieval systems to represent a first approximation to efficient search with maze organizing techniques applied as a superstructure.

Experiments may easily be formulated with existing subject catalogs and mechanized search systems to test the relative effectiveness of the suggested methods. If initial tests indicate further interest the engineering evaluation of mechanical devices may be carried out.

REFERENCES

1. MOORE, E.F. *The Shortest Path Through a Maze,* International Symposium on Switching Theory, Harvard University, April 1957. (To be published in the Annals of the Harvard Computation Laboratory.)
2. WILDHACK, W., STERN, J., SMITH, J. Documentation in Instrumentation, *American Documentation, 5,* 223–237 (October 1954).
3. MORTIMER TAUBE and ASSOCIATES. Storage and Retrieval of Information by Means of the Association of Ideas, *American Documentation, 6,* 1–18 (January 1955).
4. BRANTLINGER, R., JENKINS, N.P.H., MACK, C.D. *List of Subject Headings for Glass Technology,* Pittsburgh Plate Glass Co., Glass Division Research Laboratories, Creighton, Pa., 1954.

SUMMARY OF DISCUSSION

THE DISCUSSION of Area 6 opened with a presentation by the Chairman, Dr. John W.Tukey, of a consensus of panelists' views concerning the papers, which was clarified and made more explicit in later discussion. The papers generally introduced mathematical structure, but this was used for linguistic description rather than as a basis for applying known theorems. Panel members saw considerable usefulness in putting matters in a formal language, even though there may have been nothing new said. (It was emphasized in discussion that, for example, the clarity of statement of Mooers' paper, as illustrated by the separation of prescription and delineation spaces, was likely to be very valuable.)

Members of the panel emphasized that additional formal structures are likely to be required for any adequate theory. As mathematics should be shaped to the needs of the use to which it is put, no single mathematical technique available today, such as partial order or lattice theory, is likely of itself to carry us far in a field with such broad ramifications as that of information retrieval. After the problems of information retrieval has been carefully studied for a long time, there may be a mathematical technique appropriate to the field. But if so, it probably will not be one of the techniques known today. The panelists found the discussion of relationships among terms or descriptors good so far as it went, but pointed out that adequate accounts will have to treat more diverse and complex relations. (There was some discussion of the extent to which interfixes are analogous to, but somewhat more flexible than, Dr. Vickery's interlocking relations.)

The paper in Area 2, by Lykoudis, Liley, and Touloukian, was cited as an example of a successful combination of formulation, mathematical development, and experimental verification on a modest scale.

Information retrieval potentially involves very large systems, and analogies with other large systems might be helpful. Mechanization and automation of such systems has not necessarily reduced the complexity of functional separations that existed in the manual system. Thus, it was pointed out that the U.S. telephone system has been called the largest computer in the world, yet mechanization has not eliminated local central offices, toll-tandem offices, and the like. Instead, the long-distance call of tomorrow, like that of today, might go through a half-dozen or more exchanges. Information retrieval today is a multistage process. One should not expect machines to change this.

DIAGRAMMING THE FUNCTIONS OCCURRING IN RETRIEVAL

Three diagrams were then introduced to illustrate the multistage process in information retrieval. First, a description of the conventional index-abstract-document search was used as an example of a three-stage system. The structure of an information token was displayed as consisting of three parts: the *index,* that by which the token can be selected; the *brief,* that which is a guide to whether the next most detailed token should be examined; and the *outdex,* that which guides the retrieval of the next more detailed token. Classical examples include a subject index to abstracts, where the subject heading serves as index, the author's name and short title of the paper serve as brief, and the page or abstract number of identification serves as outdex.

The second diagram, reproduced here as Fig. 1, attempts to set out a pattern within which the first stage of a multistage information retrieval system, whether manual or mechanized, can be fitted. The querist appears both at the top and at the bottom. At the top he generates an initial vague query, probably phrased in common or vulgar language, and a query formalizer (whether librarian or future machine), then engages in an exchange with the querist in order to come up with a proper, formal query. The suggested formalization goes back to the querist in a language which he can understand—an external

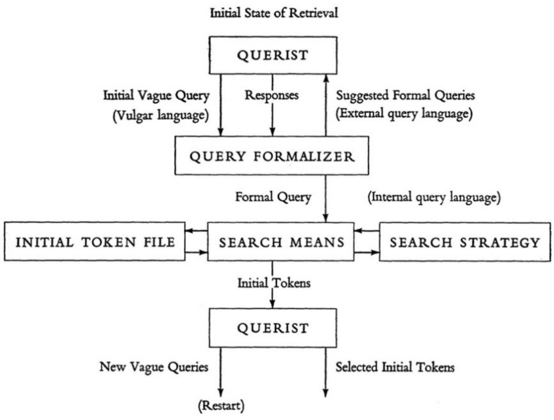


FIGURE 1.

query language. As an approved formal query goes forward into the next stage, however, it may be expressed in an internal query language, which is different both in structure and expression.

At the next level (Fig. 1) are shown a search means, a search strategy, and the file of initial information tokens. Selected initial tokens are referred to the querist who may pose new vague queries, or who may decide to examine indicated tokens of the next more detailed level, or both.

The third diagram provided for the generalized pattern of any later stage of retrieval, in which tokens of one stage are selected by the querist and are then used in a retrieval operation to select tokens of the next stage.

Discussion of this material led to apparent agreement: (1) that in this field, of information retrieval, theory is often still trying to describe library practice (Mooers); (2) that in presenting the graphical material the panelists were concerned with pedagogy and not with announcing brave new insights (Fairthorne, Tukey) but that explicit statement is useful (Touloukian); and (3) that the diagrams did not give enough attention to modification of both querist and search strategy during a search (Hayes, Mooers, Shaw).

SOME MATTERS OF THEORY

The Chairman next raised the question of how general a retrieval theory ought to be this year and next. He suggested, first, that the broad structure in which a general theory should operate should encompass all present systems, whether manual or mechanized, as well as those that appear reasonably possible in the future. Secondly, the broad structure should try to establish clear functional parts where appropriate. Detailed theories of such functional parts should then be sought. Finally, it will be important to learn by experiment how, *in functional terms,* classical retrieval is carried out.

He suggested that another important analogy could be drawn between the telephone system and information retrieval, namely, that information theory has not taught us how to build telephone systems nor what restraints we must suffer in such building. It has only taught us how well we can use a transmission channel, which is an important *functional part* of a telephone system. Information and communication theory *could* tell us how to pack (code) information that is of an anticipated nature. Information theory therefore could be expected to apply to the expression, but not the structure, of the *internal* query-language; to the organization of the search means and the initial token file and of their intercommunication; and to the organization of retrieval means, of *k*th token files, and of their intercommunication; but *not* to information retrieval as a whole.

Application of information theory did *not* seem at all likely for more difficult problems such as the structure of the internal query language, the nature and mode of application of the search strategy, or the structure and expression of the external query language. It is therefore in such areas that the hardest and most creative thought is needed.

As a final point, Dr. Tukey stressed that any sound retrieval system must learn, must modify itself steadily. It must do this, not only during the course of a difficult search but over longer periods. To this end, it must be concerned, for example, with what is asked for together, with what users finally accept together, with what users state to be related, and with what new concepts of importance are arising. As the bulk and diversity of the information to be handled continue to increase, the learning ability of retrieval systems seems likely to become one of their most important features.

Discussion emphasized convergence on two points: that the engineering of the retrieval process itself, including the organization of the file, is both of vital importance and a major intellectual challenge (Taube, Hayes, Bar-Hillel); and that organization of the tokens before transmission to the querist is often crucial (Farradane, Hayes, Fairthorne).

PRESENTATION BY PANEL MEMBERS

Mr. Russell A.Kirsch began by relating two of the Area 6 papers to the second of the three fields—indexing, coding, and machines—into which Dr. Jonker had just proposed dividing the over-all field of storage and retrieval. The Mooers paper attempted to relate three apparently quite different systems for encoding indexed information to one common formal structure. The Fairthorne paper approached the problem in a different way. In coding indexed items, one of the things that significantly affected the list of mechanisms that can manipulate the information is the language in which the information is expressed. Mr. Fairthorne appeared to be concerned with the introduction of a formalism which would enable us to handle the different coding techniques that are used in a retrieval system. Mr. Kirsch remarked that, contrary to the remarks of Professor Oettinger in the Area 5 discussions, he felt that the introduction by Harris, of the mathematical structure of groups of transformations, in his paper on linguistic transformations, was appropriate and wise, as it opened up the possibility of putting much formal knowledge to work.

Mr. Kirsch remarked concerning the apparent relative neglect, both in the papers and in the discussions, of three problems of considerable importance and difficulty. First is the formidable problem of indexing—the problem of producing from documents some type of surrogates that can be used for subsequent manipulation. A second critical problem area is that of attempting to handle natural language in a machine. Noting that there are only the two papers on this subject, one by Harris and one by Yngve, he suggested that the distribution of effort in this area is out of balance with the importance of the problem. The third problem is that of handling graphical configurations— printed characters, pictures, and diagrams.

Mr. Kirsch suggested that a principal basis for the difficulties encountered in handling parts of the retrieval problem is the tremendously diverse nature of the information. One of the reasons that we are looking for a theory is simply the matter of economy. If a theory can be developed, this would result in a considerable simplification of the problem. He called attention to the last sentence of Mr. Fairthorne's paper in Area 6: “Systems founded on neat and economical theory are neat and economical in other ways.”

Dr. Marvin L.Minsky emphasized that a modern computer can perform tasks far beyond the scope suggested to us by the nature of the individual operations, the ability to be programmed to alter its own program. Dr. Minsky began by stressing the difference between what a machine can do and what a program can do, and stating that the only limit to the latter was the cleverness of the programmer. (Once the principles and general structure of a program are worked out, they can be implemented on any general purpose computer with access to sufficient memory.)

He went on to discuss some recent developments in heuristic programming, especially as applied to programs for playing chess and checkers, and for proving theorems. In such programs, the machine is instructed what to try *first,* and how to use results of the trials to modify further action. The programmer himself may not know what trials will work, nor how to find them explicitly. Dr. Minsky estimated that in a few years a large fraction of computer time would be spent in executing such programs.

The complexity of programs built up by the action of other programs, and the inability to tell in advance what details of a heuristic program would be executed in a specific application, make the development of a detailed theory of the behavior of such programs almost impossible. Indeed, nothing so complicated has ever been successfully treated in detail by mathematics. There must be a theory of building systems without a detailed theory. This will be particularly true for a retrieval system that is to grow in sophistication.

Dr. Minsky said that an art of constructing these dreadfully complicated programs is being developed so that they can be modified without understanding them over again. We are trying to gain an understanding, perhaps sometimes even a theory, of how systems can be built which are so stable, or so resilient, that they can be modified or improved without having to be redesigned or reanalyzed as wholes.

Potential applicability to the multistage retrieval process is obvious. The programmer does not know in advance what path his program will take as it goes through a particular search. It is not known in advance whether the results of the search to a certain point may require beginning again with a modified search strategy. For example, a particular type of information token might be used, one which, if selected, would indicate not that a particular piece of information would be available for retrieval but rather that a poor search prescription had been used. It is possible to give advice to a "machine" that is able to accept such tokens. Thus, the possibility of designing really flexible systems is promised.

Dr. Minsky said that the best collection of papers on artificial intelligence and on programs to solve problems for which ways of solution were unknown would be the papers of the conference on Mechanization of Thought Processes, to be held at Teddington, England (November 25, 1958).

In later discussion, Dr. Minsky went on to say that the proper language to describe processes that involve so much returning and self-modification is not the language of classical mathematics but the language of programming. This is not the language of the 100 or so relatively trivial basic machine instructions, but rather the language embracing the far more profound operations used in automatic compilers and the like. This is a language that psychology, sociology, engineering, and retrieval have lacked. It has a further great advantage; in order to see what may happen in a complex process, one need only load the machine and press the start button.

Mr. Kirsch remarked that flow charts for a retrieval problem for the Patent Office were fantastically complicated, and that the ability to make systems without the need of detailed systems design seemed essential to him also.

Dr. Benoit Mandelbrot urged that the study of taxonomic trees, especially those developed in botany and zoology, is pertinent to the analysis of the information retrieval problem. He stressed that taxonomy is the original classification procedure where there are a large number of items having a large number of properties, only some of which are used for classificatory purposes. Taxonomy has been marked by the use of intermediate categories whereby individual species were not necessarily described directly but were grouped into families and genera that could be defined more systemically.

This fact suggested the possibility of experimentally determining how the procedures that have been followed by taxonomists would compare with the procedures which a classification specialist of today would follow if he were presented the same problem without a prior background of knowledge. Would different persons starting from different principles of classification in fact arrive at similar results?

An important reason for concern about classification is the very rapid exponential increase both in the number of items and in the number of categories that must be dealt with in a classificatory operation. Exactly such an exponential increase is encountered in the models that biologists have proposed to explain the existing taxonomies. Under circumstances of such increase, preconceived ideas of an optimal basis for classification soon collapse into disordered systems, or into something quite different from the original criteria. For example, when taxonomies of families of animals or vegetables are examined, we find that some of the statistical properties are the same in all cases and that the schemes are very far from being optimum from the standpoint of even the least sophisticated coding theory. Yet these schemes have properties perhaps even more systematic than that of dividing each sub-set into two equally probable sub-sub-sets. Thus, investigation of the growth of existing taxonomies whose branches were split apparently randomly might lead to better understanding of methods adaptable to situations of continuing growth, than might investigation of classificatory procedures constructed by splitting into equal, and equiprobable, parts as might be suggested by communication theory.

Dr. W.J.Turanski discussed three functions of a theory: it should indicate its own scope, it should describe usefully, and it should help with engineering. Just as mathematics is what mathematicians, who are identified by an attitude and habit of mind, do, so information retrieval is the retrieval, as opposed to the resynthesis, of information. Storage of basic elements alone, rather than information, would not be information retrieval. Thus, the storage of axioms, rules of procedure, and heuristics, and producing a source of theorems, would not be information retrieval.

One of the dangers of various branches of mathematics is the power of their expressive vocabulary, which is so great as to be able to describe almost anything. Not all such descriptions are interesting. Finding good modes of description which present this essential feature of interest will be important for a theory of information retrieval. It will be important to recall "Fairthorne's principle" as stated in his paper in Area 6, "…the properties of the way in which we talk about things can distort or obscure the properties of the things talked about."

A theory can help solve engineering problems. These are problems of getting things done subject to such constraints as limitations of time, money, and equipment. In this connection the facts that the population of requests may often be segregated into subpopulations and that these subpopulations change with time are often important. The latter makes inflexibility serious; the former makes retrieval networks (where the basically straight-line path of Fig. 1 is replaced by interweaving alternatives) attractive.

The discussion at this point was diverse. Miss Margaret Masterman felt that it was important to go far further into the absolutely fundamental question of what it was like for mathematics to be applied to a theory of this sort, and how one may know that he is proceeding properly in making an application. Dr. Minsky replied to the effect that just because some of the panel members are mathematicians it was not their job to say that a theory of retrieval is identical with applying mathematical theories to retrieval. Dr. Gilbert W.King urged the importance of the “breeding” of ideas and their cross-indexing in a lattice. Dr. Don Swanson raised the question of the importance of consistency from human to human in the execution of a task, as a guide to its mechanizability. Dr. Minsky stated that, since human consistencies can be due to common error or superstition, and human inconsistencies can be due to differences in background, he felt that observable consistency is not a useful criterion for mechanizability.

FURTHER PANEL PRESENTATIONS

Dr. Yehoshua Bar-Hillel opened the discussion of organization by discussing concepts of “relatedness” and “degree of relevance” which needed to be carefully distinguished. Much of his discussion was formulated in terms of documents, queries, and index terms. The elementary approach of matching the set of terms for a document against the set of terms for a query runs into many difficulties. The introduction of interfixes and other relations was made necessary by some of these difficulties. The nature of other difficulties was such as to force any theory into probabilistic terms, thus greatly complicating its structure. He felt that the best that can be done with such techniques is probably a compromise, recomplicating a simple descriptor system in the direction of a subject-heading system.

He then went on to discuss notions of relatedness of the form “A is related to B if and only if A and B occur together in some C,” and those which derive from these when chains of such relatedness of some fixed length are considered as defining a weaker notion of relatedness. The notion of relatedness of documents when they have an index term in common, and the notion of relatedness of terms when they apply together to at least one document, belong to this class. They have been tried; a formal theory exists which can be directly applied; Dr. Bar-Hillel's strong hunch was that this would not lead to any interesting results. Another notion is that of considering documents as related when requested together. If applied vigorously this would be most dangerous because it would canalize research. The notion of relatedness of indexes that appear in the same query also belongs to the same group.

The relatedness of topics and concepts is an infinitely more difficult problem than those just discussed. Dr. Bar-Hillel believed that any simple-minded approach to this problem is sure to fail. Merely going back to the indexes as such would fail. Although machines deal with terms, not with topics or concepts, the problem of relatedness of topics and concepts is basic to all these problems of relatedness.

Dr. Julian Bigelow commented on the need for explicit interpretations in describing models and the need for growing theory from the basis of a real knowledge of present-day practice. He felt that the six papers in this Area are excellent examples of model studies, and that more work along such lines would be a great contribution. In theoretical papers giving mathematical models it is most desirable to attach an appendix which gives as clear interpretations of the mathematical entities as possible. Thus he found the use of “spaces,” without an indication of whether or not ideas of “neighborhood” (and which, if any) were to be used, likely to lead to misapprehension. Neighborhoods in retrieval spaces were likely to be inhomogeneous and very peculiar.

He felt an enormous need for understanding what the empirical process is as people do it (and not as they say they do it). This would require detailed recording, in an impartial way, of many steps of communication during many inquiries. To continue adjustment of understanding and theory, such recording would need to continue through the stages of mechanization. In particular, persons entering the system should begin by describing themselves and their motivation as well as stating their formal query. He also felt that such description might be important for fully mechanized systems.

The understanding gained from full recording could serve several purposes. Near-future mechanized systems must work with people with today's habits. Adequate schemes of relatedness need guidance and validation in terms of human practice. The most profitable and powerful way to develop theory-would be to build up slowly on an empirical foundation.

Dr. Donald M.MacKay felt that problems of meaning should not be avoided. Information theory had avoided them by presupposing an extent of statistical knowledge impractical in information retrieval. Meaning for retrieved information might be definable through the needs it meets, by its selective effects on the actions or states of readiness for action of the recipients. A classification scheme for needs or states of demand for the selective function of information then would be required. Such a scheme would allow Dr. Bigelow's inquirers to describe themselves in effective terms.

He felt that the possible symmetry between the information and queries deserves further exploitation. Documents could be made active to scan need lists, at the same time that need lists are scanning documents. Syndicating together blocks of related needs could serve efficiency, just as does syndicating together blocks of related documents. The degree to which departure is made from symmetry of activity, in one direction or the other, should depend on the statistics of the situation. Infrequently used documents should scan; frequently used documents should be scanned.

Dr. I.J.Good spoke about the problems of definition and the possibility of useful definitions of clumps. The description of required information is analogous to the definition of a word. John Wisdom (The London School of Economics and Political Science, University of London) has considered the definition of a cow, which typically has many properties, no one of which is essential to its cowness. Formulation of any definition in terms of a function of qualities or of probabilities of having qualities leads to dependence on such things as time and context, and these in turn depend on others. To produce accurate definitions is very difficult.

Dr. Good went on to consider networks connecting vertices, which might represent documents, or propositions, or described customers, each connected by paths characterized by two impedances, one in each direction. Each impedance could be taken to be the reciprocal of the relevance of one point to the other. Once these relevances are numerically given one can try various rules for identifying clumps of points, clumps of clumps, clumps of clumps of clumps, and so on. These would form tree-like structures intersecting one another, and seem likely to lead to more general structures than lattices. Dr. Good did not wish to bore the audience with discussion of trials of detailed rules for identifying clumps. He did discuss the possibility of using functions of relative frequencies of words as measures of relevance. Dr. Bar-Hillel commented on the advantages of using ratios of relative frequencies to relative frequencies in “general English.” Dr. Good pointed out the advantages of ratios to relative frequencies for a larger context which was still not as broad as “general English.”

INFORMAL DISCUSSION

Area 6 discussion was resumed on an informal basis in the afternoon, and began with a discussion of Dr. Jonker's paper. After some misconceptions of the morning discussion had been clarified, Dr. Bar-Hillel stated that the term “continuum” was misleading (because of its implication of connectedness) and that it was probable that Dr. Jonker meant “linear order.” But Dr. Bar-Hillel felt that it was “clearly wrong” that all indexing systems can be arranged in a linear order; that Dr. Jonker had not tried to demonstrate, rather than state, this; that even if linear ordering were possible it would be amazing if “length of terms” provided a reasonable order; and that length of term depended very much upon the size of an alphabet, vocabulary, and so forth, chosen as a basis for counting. Dr. Minsky said that no one could believe that Dr. Bar-Hillel meant that it doesn't matter in practice how many bits it takes to specify a symbol which must be stored. He went on to say that he had difficulty in Dr. Jonker's paper in understanding whether “length of individual term” or “length of average retrieval prescription” was arranged along the “continuum.”

The discussion of these points continued without adequate distinction between a number of alternative thoughts: (1) that one important way in which index systems vary is roughly as described by Dr. Jonker, although precise comparisons are difficult or impossible; (2) that a precise definition could be given of *a* characteristic or coordinate of indexing systems which would establish the kind of order considered by Dr. Jonker without completely determining all the important characteristics of such systems; (3) that a precise definition could be given, so that systems near to one another on this coordinate would be near to one another in all relevant characters. Apparently Dr. Jonker felt that (1) was almost as good as (2), while Dr. Bar-Hillel thought that (2) without (3) was of little use.

The subject of the treatment of interfixes and relations according to Dr. Jonker was brought up. He apparently proposed to deal with them by treating clauses as large as might be necessary as single terms.

Mr. Farradane then restressed the need for the ultimate general theory of storage and information to cover *all* stages, particularly including the initial problem of indexing and the final problem of synthesizing the output information into a coherent whole. Both of these areas require analysis in terms of human thought, and it is here that his interest lies.

DISEASES OF CATS

During the morning session, Dr. Bar-Hillel had posed the question, “Which is logically more closely related to a query on ‘diseases of cats': diseases of animals, diseases of siamese cats, diseases of ducks, life cycle of cats?” He expressed his feeling that no *theory* could answer this question. In the afternoon discussion, Dr. Wall stated that he believes the thesaurus approach would work if unit concepts were employed. Dr. Bar-Hillel said that he did not have the slightest idea whether “animals” is more closely related to “cats” than is “dogs.” Mr. Farradane felt that in dealing with things related to “diseases of cats” one could assess relatedness to “diseases” and to “cats” separately, and then combine the results. Dr. Good pointed out that if words were among the vertices of a network, then adding a word automatically increases clumpiness near the word, and the existence of a machine-recognized clump suggests the need for an associated word. Mr. Farradane went on to say that the probable faults of choice of the inquirer are important in dealing with the relatedness difficulties brought out by Dr. Bar-Hillel in connection with "diseases of cats."

Questioning by Dr. Wall brought out the fact that Dr. Bar-Hillel did not mean that a librarian who understood the meanings of the words (and who could talk to the inquirer) could not deal with the "diseases of cats," but only that there was no automatic logical relationship to select a related topic (and he doubted the existence of a practical and mechanical rule).

Mr. Vickery then made a number of points about assessing relevance from relations of sets of documents and query terms. A one-to-one match of all terms would equate query and document. Both auto-abstracting and ranking by the querist of nearby sets of terms offer promise. What is attempted in libraries is to rank the documents in a scale of relevance based on the observations of previous use and satisfied users. This is what any classification or thesaurus structure does. The scale of relevance changes from time to time and from one questioner to another. The job is to find ways of making systems more sensitive to this changing relevance; not to worry that the "relevance" is only very, very probable.

Mr. Fairthorne pointed out that the relevance of a document and the description of its "contents" are not the same. He went on to say that one reason why biological types of classification, as suggested by Dr. Mandelbrot, are not used by librarians is that they serve to pull out extinct animals; librarians are not concerned with pulling out extinct books. He emphasized that a pair of documents once irrelevant can become relevant, but not vice versa. As a consequence there was trouble with the denial of the "excluded middle." Such strong symmetric algebras as lattice theory seemed to him inappropriate. He concluded by emphasizing the need for two algebras, one for retrieval of "all but not only" and the other for the retrieval of "only but not all." (See *American Documentation,* July 1958, pages 159–164, for further detail.) The "backlash" between these two algebras is important.

Dr. Taube then pointed out past carelessness about the distinction between plans (or dreams) and actual operating machines, alluding to the previous day's discussion of the Perceptron. He then challenged the existence of machines that can act heuristically. Dr. Minsky pointed out that there are machines that play chess and checkers (checkers well enough to beat good amateurs, chess poorly); that there are machines which can, and do, prove theorems in the propositional calculus; and that there are at least four or five separate instances of general purpose computers being used for character recognition.

FANO'S DREAM

Dr. Robert M.Fano presented his dream of a library service. He said that he would be unwilling to phrase his questions in any prescribed words (descriptors, classifications, or others); his thoughts did not fit in such paths; he was individualistic in that matter. He would be prepared to list documents which were "like" what he requested, and to list others which were "unlike." (This type of query language had already been brought up during Dr. Bigelow's presentation.)

Dr. Fano envisioned such a retrieval technique embedded in a library system with a closed-circuit TV and keyboard in his office, so that he could request from his chair, and either read what was retrieved on the screen or have copies sent to him. (The screen would also serve for verbal messages from the retrieval computer as the search progressed.)

How could such service be accomplished? He felt that storage of one number for each pair of related documents would suffice. He preferred not to think of these numbers as measures of relevance (as in Dr. Good's presentation) but as experimentally obtained measures of association, with useful operational properties. The basic problem in using them was to obtain the effect of searching through the library of documents without performing a correspondingly large number of operations. He had some definite ideas as to how this might be done, but this discussion session was neither the place nor the time to give details. In any case, such ideas could be only a starting point for an experimental investigation of the problem.

Dr. Fano felt that such a library system would be a very large system that could not be designed a priori, but only developed in successive steps through experimentation. Its operation could not be tried out with pencil and paper or on a small library. He made the point very strongly that extrapolating results from a small system to a large one is a very dangerous business. The major technical difficulty would not be the availability of suitable digital computers in view of the fantastic progress in the development of micro-components. The bottleneck would be in programming the computer. He agreed with Dr. Minsky that quite a bit of progress is being made, and that the problems are very tough but not unsolvable.

How could suitable measures of association be developed? He thought that citations of one document by another would be a web strong enough to establish a beginning. Progress (he did not like to say learning) could then be pro grammed by calling for the modifying of associations in terms of such observations as frequency of co-use of two documents by one requester.

Costs of the final system would not be prohibitive, although large. Development would be expensive, as it always is. He stressed that the whole field needs much more experimental research (not to be confused with development of hardware) and that funds for such a costly resaerch have been lacking.

Discussion of Fano's dream began with comments from the chair that (1) associative systems have the advantage of no classifications to wear out and (2) are likely to take as unit documents portions of present day documents. Mr. Taffee Tanimoto (IBM) and Mr. Williams (duPont) stated that they had programmed (independently) and used at least the first phases of the associative access system dreamed by Dr. Fano. Mr. Tanimoto's system was concerned with the classification of plants. Mr. Williams emphasized that the approximation in indexing what was written was only added to the approximation of what was written to the writer's thoughts. Dr. Bar-Hillel pointed out that Dr. Fano's dream was dreamed by a Russian Academician between 1947 and 1951.

Dr. Bar-Hillel went on to state that while Dr. Fano's dream could be realized, the question is whether it should be. Large amounts of money would be required, and one should be sure that this would be a good way to spend it. There are many schemes that could be tried out, but only those for which strong plausibility arguments could be given deserve trial.

Dr. Bar-Hillel was sorry that Dr. Fano had not had time to be explicit about his suggested measures of relatedness. Dr. Bar-Hillel would be ready, within an hour after being given a measure, to present another measure which is better in a well defined sense. He felt, for instance, that being co-quoted, which is based on knowledge, is better than being co-requested, which is based on ignorance. But this, too, could be improved on.

The Chairman suggested that Dr. Bar-Hillel be hired for 200 hours to make 100 successive improvements. Dr. Fano suggested that it would be better for those interested to get together and “improve” measures jointly. Dr. Fano went on to emphasize the inadequacy of persuasive reasons without experiment. It takes the combination to produce a rapid and natural evolution of concepts and knowledge. Mr. Moorers took Dr. Fano's dream quite seriously. Just as computers had buffer units, to mediate between machines, the dream needed a duffer unit, a simple and fool-proof mechanism with keyboard and printout, to mediate between people and machines. Mr. Norman Ball (National Science Foundation) called Dr. Fano's attention to a very large-scale operating manual (non-mechanical) model of his dream called the Patent

Office Search Room. Dr. Fano confessed that his dream represents the ultimization of what he believes goes on in any library. Mr. Hayes emphasized the problem of storage and pulling out in a large library, and argued that the purpose of classification is more fundamentally that of easy access than that of determining which document is wanted. Dr. Baumann (M.I.T.) asked the audience to think about the “human being,” with the capability to read all the books in all the libraries, to retain what he had read, and to interpret what he had retained. This would be a long step toward Dr. Fano's dream. If this capability were ever to be realized by an intelligent machine, there might be need for care that the machine did not become super-intelligent and do our thinking for us.

A librarians comment was made by Mr. Cleverdon. From the discussion which had gone on all day, it was quite obvious to him that the mathematicians are trying to find a mathematical statement of what librarians have been doing for a long time. This has tended to cause the librarians a bit of heat, because they did not realize why this was suddenly the case. Mr. Fairthorne's 1947 paper on mathematics of classification was an attempt to help librarians with their work. Today's activities are different, probably because of the possibility of putting some library operations onto machines. If the librarians' activities are to be mechanized, someone must find exactly how and why librarians do everything they do. And to find out “why” librarians do what they do, experiments are needed. He had one in progress, but others are needed. A moderate percentage of the total investment in information retrieval should be used for experiments; all of it should not be used for machines.

JOHN W.TUKEY, *Rapporteur and Discussion Panel Chairman*
LAWRENCE F.BUCKLAND, *Area Program Chairman*

AREA 7

RESPONSIBILITIES OF GOVERNMENT, PROFESSIONAL SOCIETIES, UNIVERSITIES, AND INDUSTRY FOR IMPROVED INFORMATION SERVICES AND RESEARCH

AREA ORGANIZATION

Authors of Papers

Members of Discussion Panel

Chairman: VERNER W.CLAPP, Council on Library Resources, Inc., Washington, D.C.

B.W.ADKINSON, Office of Scientific Information, National Science Foundation, Washington, D.C.

W.O.BAKER, Bell Telephone Laboratories, Murray Hill, N.J.

NOBLE CLARK, Agricultural Experiment Station, The University of Wisconsin, Madison, Wis.

SIR HERBERT HOWARD, Commonwealth Agricultural Bureaux, Farnham Royal, Bucks, England

D.C.MARTIN, The Royal Society, London, England

J.C.SECONDI, Provincial Health Service, Ministry of Public Health, Buenos Aires, Argentina

JESSE SHERA, School of Library Science, Western Reserve University, Cleveland, Ohio

C.E.SUNDERLIN, European Research Associates, Brussels, Belgium

RAYMUND L.ZWEMER, U.S. Department of State, Washington, D.C.

PROPOSED SCOPE OF AREA 7

THE TASK OF developing and maintaining effective scientific information services is of such magnitude as to require the resources and cooperation of all organizations with an interest in the progress of science. To varying degrees, governments throughout the world have undertaken to provide, or contribute support to, information centers and services; professional societies and industrial organizations have been active in fostering interest in and support of documentation work in their respective fields. A study of these various programs and the results achieved should furnish important clues as to the other problems involved, the areas in which greater attention is most desirable, the effects of financial support on the character of the services, and the directions in which further probing would seem to offer the highest dividends.[1]

Compared to the level of production of new scientific information, research on the organization and use of scientific information is now very limited. The belief is widely held that whatever may be the current level of support for scientific information services as such, significant improvements in their effectiveness are not likely to be achieved unless a vastly increased amount of research and development is undertaken in these fields. Accordingly, the following questions need to be asked: If, because of the size of the problem, we must look to national governments predominantly for support, what residual responsibilities remain with the professional societies? Should they not be expected to contribute by identifying the scientific and professional problems (and needs) involved, and to suggest avenues to their solution? Although the professional societies are necessarily concerned with current publication of the results of current research, should they not also be more concerned than at present with assuring the availability of the past literature through the tools of retrospective search? In addition, what is needed to provide unity in documentation research and to overcome present scattering of effort? How can the individual operations of publication, indexing, microfilming, and so forth be brought together and recognized as parts of a whole? How can the professional societies in the various subject disciplines contribute special insights into special problems which may well illuminate fundamental documentational principles in still other disciplines?

[1] The proposed scope of the Conference Area, as shown here, was prepared during the Spring and Summer of 1956 and provided to all potential contributors as a guide to the aims of the Conference.

In addition to the problems relating to the support of effective scientific information services, and to the problems related to research and development regarding such services, a third problem area may be identified. This is the area of training for activity in scientific documentation work—either in the production of services themselves, or in research to promote their efficiency. We need a survey of the training facilities now available as a basis for calculating the additional facilities required to meet current and future needs.

SUGGESTIONS FOR CONFERENCE PAPERS

1. Differences in various national and international arrangements for financial support of science information services: relationship of the supporting organizations to such services.
2. Conditions and methods necessary for promoting research on organization and use of scientific information and on the problems of scientific documentation.
3. Implications, for training, of the requirements of the science information services and of the needs for research in the problems of scientific documentation.

Responsibilities for Scientific Information in Biology: Proposal for Financing a Comprehensive System

MILTON O.LEE

This paper will deal with responsibilities for information storage-retrieval and with mechanisms for discharge of these responsibilities in the field of biological science only. It is perhaps applicable to other fields of natural science as well, but the author is not as acutely aware of the particular situations and problems in other fields, or of the mechanisms now serving them. Also, the circumstances in biology are perhaps more desperate than in the other large areas of science.

Biology is the most extensive field of natural science. Its core of basic disciplines, each rather poorly delimited, is large, ranging from the numerous divisions of plant science, animal science, microbiology, and bacteriology to human biology. It embraces the sciences of genetics, evolution, morphology, taxonomy, biochemistry, biophysics, physiology, and ecology. From another point of view it ranges from the molecular level through the tissue, organ, regulatory, and organism levels to the ecological, behavioral, and social levels.

In addition, the applications of biological science overbalance the basic core sciences. These applications are chiefly, but not exclusively, in the technologies of agriculture, medicine, veterinary medicine, dentistry, and public health. These applications in the technologies have become so necessary and vital in our civilization that they have overshadowed the basic life sciences upon which they depend. In the number of professional people employed, in the amount of research work undertaken, and in the financial support for such research, the applied fields of biology far surpass the basic biological sciences themselves.

With such a great diversity of both applications and basic disciplines, the communication of new information, its storage, and retrospective search and

MILTON O.LEE The American Physiological Society, Washington, D.C.

retrieval have posed enormous problems, probably much greater than in any other of the natural sciences. The mechanisms presently in operation are inadequate in varying degree, both with regard to the particular service each attempts to render and in the coordination and interrelations of each service with the others.

Responsibilities for handling various aspects of information and communication problems have been assumed by necessity-conscious groups, and organizations have been set up largely on an *ad hoc* basis and without much consideration of overall relations or total adequacy. It may be useful at this time to attempt to delimit the responsibilities that are now tacitly assumed by or that might logically be assigned to the several elements that are concerned with biological information. For the purposes of this paper, these elements are considered to be:

1. The research funds and the agencies, public and private, that provide financial support for the production and utilization of new information.
2. The scientist-scholars who produce the information.
3. The media (mostly journals) that disseminate the new information initially.
4. Information services and centers: general abstracting and indexing media for classification and storage-retrieval, similar services in specialized areas, data processing centers, and compilations such as reviews, bibliographies, handbooks, and compendia.

The analysis and the proposal that are made are with reference to the United States only. However, many of the delineations of responsibility are probably applicable generally, and the mechanism suggested for meeting responsibility may be applicable to the situation in other countries.

THE RESPONSIBILITIES OF RESEARCH FUNDS AND AGENCIES SUPPORTING RESEARCH

More than 60 percent of the funds that support biological research in the United States at present come from research-project grants. These grants are for budgeted expenditures for specific projects, planned by the researcher or research team and approved by the fund-granting agency. The pattern of the support is so well known that it is only necessary to summarize. The budgets for such planned projects may include salaries of investigators, technical and clerical help, necessary instruments and apparatus, supplies, necessary travel, occasionally the construction of facilities, and an institutional overhead charge.

Seldom is there any specific provision in budgets for the publication of the results of the investigations and probably never for any services of storage-retrieval. Yet the funds expended for the researches are almost entirely wasted unless the results are communicated and, even if published, will after a period of years certainly have been wasted to the extent that retrieval by future workers becomes extremely difficult or impossible in a practical sense.

Should we not clearly recognize and accept the full responsibility of research funds and research budgets to provide financial support not only for primary communication as an essential part of investigation but also for the full job of permanent storage-retrieval? The cost of meeting adequately the total responsibility would be small in proportion to our total research expenditure— of the order of less than 5 percent. Some small but encouraging beginnings in the acceptance of such responsibility have been made, notably in the field of physics in the United States, for the support of primary publication. Extension to and acceptance by supporting funds of the direct responsibility for storage-retrieval is the final and seemingly inevitable step.

RESPONSIBILITIES OF THE SCIENTIST, RESEARCHER, AND TECHNOLOGIST

The production of new biological information has been traditionally the privilege and responsibility of the scientists, researchers, and technologists who have chosen biological science or applied biology for their professional careers. Biologists are located mostly in our universities and colleges where they also produce the new generation of their kind, in institutes and foundations devoted to biological research, in various local, state, and federal government institutions, and, to some extent, in industry, agricultural experiment stations, hospitals, clinical centers, and health centers.

Biologists have always been aware of their responsibility, but their awareness has been sharpened and intensified by the increasing practice over the last twenty-five years of dependence upon specific project-research grants for support of their studies. This method of support has become a standard pattern and carries some implications that are deplored by some but which nevertheless are real and weighty.

Since new information is the only product resulting from the expenditure of these vast research funds, researchers feel it is necessary to communicate periodically their research results, whether or not they are fragmentary and incomplete in giving answers to the large questions posed by the projects. Reporting, i.e., publication, is indeed often necessary to secure continuation of support. Such reporting of fragments may serve to accelerate the overall progress of biological science over a period of time, but its increasing practice poses acute problems for the journals of primary publication and particularly for the exist ing systems for storage, search, and retrieval of the information. Traditionally, the responsibility of the researcher ended when he secured acceptance for publication of his research results in an appropriate journal. If he became an authority in his area, his privilege and responsibility extended to the authorship of occasional scholarly reviews, chapters in textbooks and handbooks and in secondary compendia of other sorts.

Have we not arrived at the time when our concept of the responsibility of the author of new information must be broadened? Should he be responsible not only for the publication, but also for its tagging, labeling, identification in one way or another to keep it permanently available in what might be called our “scientific consciousness?” Should not the author take as much of the responsibility for the intellectual part of the whole job of communication as it is practicable for him to do? His intellectual contribution is the *sine qua non* for the production of new scientific knowledge. It is certainly to his advantage to ensure that retrieval of his contribution of information will be possible for other workers without extreme difficulty for at least another generation. There is reason to believe that he will accept willingly the extension of his intellectual effort to the production of a suitable abstract of his communication, indicia for indexing and classification, and even to supply upon request specific information for data-processing systems. It should not be forgotten that the cost of the intellectual effort, if done at second hand, is generally the most expensive part of any system for information storage-retrieval. Part of it at least can be done more efficiently by authors than by others.

RESPONSIBILITIES OF JOURNALS (OF PRIMARY PUBLICATION)

As new biological information is discovered, it is communicated and disseminated chiefly through societies and the journals which they sponsor and control editorially. Indeed, communication in one form or another seems to be the main or sole function of most biological societies. The job of communication itself is done much less than perfectly on the whole and poorly in many cases. Biology has lacked the cohesiveness that has characterized chemistry and physics, and has splintered into a large number of small societies narrow in their specialized fields of interest. New ones are continually being formed. Their journals of primary publication serve rather narrow areas of interest and have small subscription lists. Circulations of more than 5000 are unusual; circulations of 500 to 1000 are common. While membership in a society usually includes a subscription to the society's journal, the individual biologist usually receives only a very few journals (perhaps 2 to 6) and depends largely upon library services for access to published communications.

Fifty to sixty years ago, when some of our present biological societies were being formed, the need for secondary media of communication barely existed. While there were occasional grumblings over the difficulty of getting access to papers in obscure small-circulation journals and concern over the increasing amount of material, by and large the individual biologist could read all the literature that was currently produced in his particular area of interest, still had time to locate and read many of the important biological papers outside of his special area, and could push his own research work along too. For search of the older literature he used the "handbooks," other review media, the several indexes existing such as the *Catalog of the Surgeon General's Library,* the *Concilium Bibliographicum,* and perhaps decennial subject indexes of his special journals. Also, then as now, he "pyramided" from the lists of references in current papers.

Our societies obviously considered their responsibility ended with provision for primary publication and almost never faced the then minor but burgeoning problem of information storage, search, and retrieval. When they did, belatedly, realize the situation acutely, the problems had already grown too large for them to handle.

It is not too late, however, for our societies and their primary journals to examine anew their functions in communication and their future place and role in the light of present day conditions and needs. For some strange reason, biological journals and editors are extremely resistant to change and adaptation to changes in our scientific environment. One is tempted to predict that, like the dinosaurs, they will become extinct. For example, the proposals made by Bernal ten years ago for a radical change in our system of primary publication, attractive and desirable in principle as they are, have met with general resistance and opposition. Hence, one must assume that in the foreseeable future our mechanisms for primary publication will continue in about their present form.

However, there is reason to believe that our journals would be willing to gradually broaden their responsibilities to cover search and storage-retrieval aspects, without changing their fundamental character. Acceptance of the following extensions of responsibility by several hundred important biological journals that publish basic research papers would seem to be possible:

1. The responsibility for furnishing a suitable abstract and indicia of every research paper published to the secondary service medium of choice for abstracting and indexing, and to such other general systems for information retrieval as may be developed to serve biologists. Supply of abstracts and index terms to the abstracting medium of choice would normally be done at the stage of page proof when the ultimate journal pagination is set in the citation.

2. Editorial responsibility for rigorous review and quality control of authorfurnished abstracts and indicia.
3. Publication of an abstract, in lieu of a summary, in a standardized form at the beginning of each paper in the journal; and free grant of permission to reproduce the abstract by any specialized information service.
4. Selection of a single common abstracting service, at least within a country.

RESPONSIBILITIES OF ABSTRACTING AND INDEXING MEDIA

There is no need in this paper to analyze in full or in detail the responsibilities of our numerous abstracting and indexing services for information storage and retrieval. However, the assumption of the responsibilities by authors and journals as outlined above demands some changes in procedure by a national abstracting service such as *Biological Abstracts,* and also offers eventually the possibility of more prompt publication, more complete coverage, and considerable reduction of the cost per abstract.

1. Give priority to and publish promptly the abstracts furnished in pageproof form by the cooperating journals.
2. Go as rapidly as possible to direct photo-offset reproduction of the printed abstracts supplied by journal editors. The preparation of *Biological Abstracts,* for example, would thus change from one of typewriter-composition, careful proofreading, correction, and arranging to a simple cut-arrange-paste job ready for photographic reduction and plate making.
3. As soon as a sufficient number of basic biological journals within a country are cooperating in supplying abstracts in page proof, restrict the abstracting medium to those and to others as they join in the cooperative enterprise. This means that papers published in non-cooperating journals would be excluded from the retrieval mechanisms of the large national abstract and index service. The pressures generated would probably soon compel cooperation by such journals or else authors who desired more than ephemeral attention to their papers would desert them.
4. The number of cooperating journals should be expanded as rapidly as possible, with the aim that at least all the important ones in both basic and applied fields of biology would be included.

When these steps are accomplished, prompt appearance of abstracts would be made easy, indeed almost automatic—often they might appear in the abstract service before the article appeared in the primary journal. The cost per abstract of publishing such a medium as *Biological Abstracts* would be decreased substantially by elimination of composition, proofreading, and correction. Savings could be used for improvement of the indexes, for extending the cover age of foreign journals, and for the development of additional search and retrieval services that are needed. Whether or not authors can supply indicia items sufficiently good that professional indexers can quickly transform them into standard entries is questionable, but the experiment should be made. If it is found that they are indeed educable, the job of indexing would be facilitated and made less costly because the intellectual effort required would to a large extent be supplied by the author. In either case the author would benefit from some acquaintance with indexing.

Abstracts and indices may prove eventually not to be our best markers for retrieval of biological information. They have become firmly established mechanisms at the present time, however, and it seems logical to push them to the limit of their capabilities rather than to try to supplant them. However, new systems and mechanisms for classification, data processing, search and storage-retrieval of information, and the extension of abstracting and bibliographic services in specialized fields should be put into operation as rapidly as these coincide with demonstrated need.

RESPONSIBILITIES OF A NATIONAL BIOLOGICAL INFORMATION CENTER

There is increasing awareness of the need for coordination and adequate financial support for a number of our present individual and uncoordinated activities whose purposes are to make biological information easily available in usable form. Some of these in the United States are the *Chemical-Biological Coordination Center,* the *Handbook of Biological Data,* the *Medical-Sciences Information Exchange,* various compendia and bibliographies in specialized fields. The sad experience of some of these in securing adequate funds over a period of years has caused abandonment or curtailment of the unfinished project in some cases and an inhibition of efforts in starting other desirable projects. A national information center able to support, coordinate, and in some cases integrate such activities would be advantageous. Centralized operation of all activities would probably not be desirable, however.

Recently proposals have been broached for the establishment of a Science Information Center in the United States to serve and coordinate the information activities for all fields of science in this country. It is very doubtful that a science-wide enterprise with this scope and with the uncertain and variable sources of support envisaged for it could serve satisfactorily the detailed needs of biological science, urgent and presently unsatisfied though these needs are. Rather, a national center for each of the major areas of science appears preferable, even though some minor overlaps would result. Each center should be adapted to the particular needs of its field, since these differ among the areas of physics, chemistry, and biology, and should also be adapted to the stage of development of information exchange that its area has attained.

A national Biological Information Center that could concentrate on the somewhat special and unique problems of biological information would appear to be practicable, provided that it were not organized and controlled by government, and that it had assurance of adequate financial support not only at its beginning but year after year for many years.

Such a center for the United States might well start with *Biological Abstracts* as its nucleus, but should both expand and speed up its abstracting and indexing services as rapidly as cooperative arrangements can be made with biological journals. It should supply information services including but not limited to abstracts and indexes for specialized and applied areas of biology. It should be prepared to search, organize, classify, and process information and data for machine storage-retrieval.

A PROPOSAL FOR FINANCING

What is the magnitude of the cost of a comprehensive activity for storage-retrieval of biological information, together with the additional services that should be developed with it? What are the possible sources of financial support? The cost will be a continuing one, year after year indefinitely. With a modest start at perhaps one-fourth or one-fifth of its eventual level, the cost would be expected to increase as rapidly as the services were rounded and completed, requiring perhaps five years.

Only an approximate estimate can be made of the present cost of our patchwork of non-government information services and mechanisms. For biology and its applied fields in the United States, the annual expenditure is not far from $1 million. This includes the general and special-field abstracting, indexing, and bibliographic services rendered by societies and other non-profit organizations which are financed chiefly by subscriptions of users. The figure does not include the cost of primary journal publication, review journals, library services, or any information services operated and financed by agencies of the federal government.

The funds needed annually for a non-governmental enterprise for biology are of the order of $5 million at the start, with increases over a few years to a level of at least $15 million. To finance an enterprise of this magnitude through annual gifts and contributions seems manifestly impossible. Neither foundations nor industries could commit themselves to making the very sizable grants needed for a long period of years. Direct federal appropriations from public funds would be impossible except to a governmental agency for a government-controlled operation. This is about the least desirable solution to biological scientists. Support through subscriptions by users of services has not proved to be sufficient in the past nor is it likely that such support for a comprehensive activity could amount to more than one-fifth of the amount needed. It is not generally realized that the cost of information services at a really effective level greatly exceeds the cost of primary publication.

What is needed is a source of funds adequate for the activities as they are developed, as nearly automatic in supply as possible, and directly proportional to the total research expenditure as it changes from year to year.

In essence, the proposal made here is that the cost of providing an adequate information service for biology be accepted as a part of the responsibility of research support through all our agencies, public and private, that make grants of funds in support of research projects or of research programs.

In operation, it would be desirable that all large fund-granting agencies pay directly to a biological information center an amount, say 4 percent, of their annual funds for support of biological research projects and programs that will involve information storage-retrieval services. For many of the smaller fund sources, direct levy at the source would not be practicable. Here our journals of primary publication would play an essential role in levying a charge for the publication of acceptable papers that report research supported by such sources of funds. This charge, uniformly based per thousand words of the printed paper, could be collected by the journal and forwarded to the biological information center. For papers whose authors had no supporting funds in the way of grants or research budgets, the charge would be forgiven. In the cases of some journals publishing research, none of which had grant or budgeted support, no charge at all would be made.

In return for its services of supplying edited abstracts and indicia and for collection of publication charges for certain papers, each cooperating journal would receive funds from the biological information center equivalent to, say, 25 percent of the levy of 4 percent on research funds represented in the papers it published.

Our biological journals are the key mechanism in this proposal. The willingness of 100 to 200 of them to enter into a cooperative arrangement for a Biological Information Center would be necessary. The arrangement would at least partly solve the vexing financial problems that most primary publication journals now face. The automatic annual addition of funds to their individual economies equivalent to a page charge of about $22 would enable them to clear their backlogs of accepted papers and to keep abreast of future publication demands. Proposals made some twenty-five years ago that research funds subsidize primary publication through page charges were generally rejected by biologists though accepted by physicists. Now some biological journals have turned to this mechanism, and others are considering it. One obstacle, the refusal of some government agencies to pay page charges for papers from their laboratories, has now been almost entirely removed.

EXAMPLE OF A BIOLOGICAL JOURNAL

I. *Assumptions:* A journal publishing 2 volumes per year; 700 pages of text per volume or 1400 pages per year; 750 words per page (or equivalent in tables and figures); 6 pages (4500 words or equivalent) per article; 233 articles per year.

Sixty percent or 140 of the published papers were supported by grants from agencies that paid directly to a Biological Information Center 4% of their grant funds; 35% or 82 of the published papers required a direct charge of $90 per thousand words at the time of publication in the journal; 5% or 12 of the published articles were from authors lacking any supporting funds for the payment of the charge; for these 75% of the regular levy was paid by the journal to the Biological Information Center.

Average expenditure of research-grant funds represented by each paper — Ten Thousand Dollars.

II. *Levies and Allocations* with respect to the example journal.

1.	To Biological Information Center		
	a) Directly from sources of research grants. 4% of grants represented by 140 papers	$56,000	
	b) Charges made by journal on 82 papers @ S90/1M words	33,000	
	Total receipts to Biological Information Center		$89,000
2.	Allocation to the journal		
	25% of receipts of Biological Information Center	$22,200	
	Less subsidy paid by journal on 12 papers	3,600	
	Net income to journal		18,600
3.	Net income to Biological Information Center		$70,400

For 200 cooperating journals operating at this average level, the net income to Biological Information Center would be $14 million, and the total net income to the 200 journals would be $3.7 million.

OBSTACLES

The difficulties and obstacles of the proposal should be outlined as well as its advantages. Some of these are the following:

a) The difficulty and slowness are considerable of getting a sufficiently large number of journals to participate in a uniform and detailed plan.

b) General acceptance by fund-granting foundations and agencies, particularly the large ones, would be necessary.

c) General acceptance by other sources of research funds of inclusion of a 4% charge in research budgets. How could papers resulting from unbudgeted research expenditures in academic departments pay a page charge?

d) The apprehension frequently expressed that any discrimination either as to acceptance or to priority for publication would have to be dispelled through careful and appropriate provisions in a plan.

e) Should only the research expenditures that result in publishable information be taxed, or should even the research ventures that failed to produce useful information be included?

f) The amount of participation of journals, usually in an applied field, that publish only a small number of papers that should be included in an information-retrieval system would have to be determined.

g) Research papers resulting from fellowships.

h) The participation of funds that support research fellowships intended primarily to increase the scientific competence of the holders would have to be determined. Should levies on these be made at the source but at a lower rate?

Responsibility for the Development of Scientific Information as a National Resource

HAZEL MEWS

Scientific information may be considered to be a national resource, as useful in its own way as any other national resource, material or intellectual, provided it is wisely exploited and planned. Once this concept is accepted it follows that the development of this national resource is a national responsibility, although not necessarily the exclusive responsibility of the central government. In a democracy, at any rate, this responsibility will devolve on many bodies; government departments, universities, research organisations and professional societies should all play their part in co-ordinating existing sources of information, improving weak ones and initiating any necessary new services. There follow some suggestions as to the interlocking functions of these various bodies; the illustrations cited are based mainly on experience in southern Africa, partly because the pattern in a new and small country is clearer to see than in Europe and America, and partly because such illustrations may be of help to other small countries.

As a first step towards adequate planning, a survey and evaluation of existing resources and immediate needs is necessary and this should be put before a top-level body (such as a National Advisory Council on Scientific Policy) that has sufficient prestige to bring about some effective action. Such surveys, in varying degrees of completeness, were recently put before the South African Council for Scientific and Industrial Research (C.S.I.R.) in the Union of South Africa and the Prime Minister's Office in the Federation of Rhodesia and Nyasaland. In the case of the Union, the sub-committee of C.S.I.R. that gave special attention to the matter included the Chairman and one member of the newly formed Advisory Council on Scientific Policy. The planning and coordination of scientific information in the past has been frustrated to an ap

HAZEL MEWS Department of Librarianship, University of the Witwatersrand, Johannesburg, Union of South Africa.

preciable degree by the fact that most co-operation was initiated by librarians and information workers themselves and they had very limited powers to put plans into action. High level support for future plans is, therefore, extremely important. It is equally important, however, that the co-operation of the librarians and other professional workers in information services should be secured as early as possible, since the smooth working of any plans depends ultimately on them. In South Africa the C.S.I.R., immediately it had approved a five-point plan of development, sent this to the South African Library Association, expressing the hope that, with the aid of the Association, the recommendations would in due course bear fruit. In the Federation, the adviser on co-ordination and development of scientific and industrial research, who had been asked to consider specifically (along with four other points) "the organisation of scientific and technical information services," called together local librarians and research workers so that the plans could be discussed in their earliest stages of formulation.

One of the first responsibilities of the central government is the setting up of a central scientific and technical library and a national scientific and technical information service, either as an independent entity or attached to some existing institution. Unesco has recognised this need as a primary one by organising central scientific and technical documentation centres in countries it was asked to assist, for instance, Insdoc, the scientific documentation centre attached to the Indian National Physical Laboratory. In the Union of South Africa the Library and Information Division of the C.S.I.R. was from the beginning planned on broader lines than as a library service to C.S.I.R. laboratories only, and it is recognised as the central scientific library and information centre in South Africa. At the time of writing, the Federation is considering grafting a similar service onto the Library of the new Rhodesia University College or onto one of the existing scientific research institutes in the country.

Central government also has some responsibility to encourage and assist local services of a more limited kind in the main scientific centres and industrial areas, to relieve pressure on the central services and to give a rapid basic service on the spot. Whether such basic services are provided by university, research institution, public library, or technical college depends on local circumstances. In three of the main industrial areas of South Africa, such services are based respectively on the university library, the public library, and the technical college library.

The scientific societies have varied responsibilities, some old and some new; they include the publishing of specialised journals, organising or assisting abstracting services, encouraging the filling of gaps in the literature coverage of their subjects, defining their exact needs as far as information services are con cerned, and educating their members to become information-minded. Until recently few practising scientists gave much precise thought to information matters and not all of them yet realise that a good deal of thought and planning has already been applied to these problems by documentalists and other specialists. The scientific societies have an educational function in this respect. Societies similar to Aslib in Britain, the Special Libraries Association and American Documentation Institute in America, the International Federation of Documentation and the national documentation institutes on the Continent have a role to play that is complementary to that of the purely scientific and technical or the purely professional librarians' associations, since they include many different kinds of workers in the information field and can weld together varying types of experience to work on these problems and to promote understanding of needs and of possibilities.

Some universities, in spite of overcrowded curricula, might be more alive to their responsibility for giving their science and engineering students some knowledge of the basic publications in their own field. Some graduates, in South Africa at least, face their professional careers with only a textbook background as far as their “literature” is concerned. As far as the chemists are concerned, their own dependence on previous work and such influences as that of the American Chemical Society's new species of “literature chemists” is leading more chemistry departments in universities to give training in searching the literature. In South Africa the engineers are also facing up to this problem. The University of Natal's Librarian now gives a course on engineering information to Natal students, and last year the new Faculty of Engineering at the University of Pretoria asked the Head of the C.S.I.R. Library and Information Division to lecture to the staff and students on sources of engineering information in South Africa. The reciprocal responsibility of information services to respond to such calls is obvious.

Universities and technical colleges which have departments or schools of librarianship should make sure that all the teaching provided in the country should not concentrate only on the training of staff for public and university libraries, but at least one institution in each country should pay particular attention to scientific information techniques. Of the six university schools of librarianship in the Union of South Africa, the one in the most highly industrialised area, i.e., the University of the Witwatersrand in Johannesburg, is specialising in training staff for scientific and technical information services. It would also help towards good scientific information services if students at the universities were encouraged to regard such work as a worthwhile career.

Research organizations and industrial firms must be willing to spend adequate sums on their own services. Southern Africa benefits in this respect from the presence of firms with overseas connexions who have already realised in Europe or America the value of such services and who find their value enhanced in places far removed from the great centres of scientific and technical activity. Smaller firms need to be encouraged and helped by the central government to set up small services of their own and to be guided in the organisation and administration of such services. The South African C.S.I.R., as a semi-government body, felt it should give help in this way, and it published a guide to the organisation of technical libraries for South African industry. It also held seminars on technical information in the main industrial centres of the Union; these were attended with enthusiasm by local scientists, industrialists, and library staffs, and, in addition to the main purpose of the seminars, the value of the contacts brought about between such local workers of different kinds was in itself a justification of the experiment. C.S.I.R. also invites staffs from the technical libraries of research institutions and industrial firms to spend some time working in its own Information Division, so that a basic idea of what information work involves can be imparted to them, together with a bird's-eye view of the entire South African information picture as seen from the centre.

Research in information problems has passed its own sealing wax and string period, when it was undertaken as spare-time work by people on the job. It now needs to be encouraged by bursaries and grants from the bodies that give money for scientific research or by industrial firms who are particularly interested in certain information problems. The lead the National Science Foundation and the Ford Foundation have given in America could be followed by official bodies in other countries. At present, however, candidates qualified to undertake such work in the newer countries are few. Last year the South African C.S.I.R. advertised a bursary for an investigation into desirable standards of staffing, equipment, book and periodical stock, photographic services, etc., in scientific, technical, and medical libraries in the Union, but it received no firm applications: the one tentative applicant could not be spared from his work as a practising information officer.

As scientific information is now a world-wide activity, the two-way flow of information and co-operation in professional information circles could be promoted by placing at the offices of scientific attaches and scientific liaison officials in the various overseas countries, personnel trained in information work and with a sound working knowledge of the information picture back home. The South African scientific liaison officers, who, before being posted to the London British Commonwealth Scientific Office, received some elementary instruction in searching for chemical information and the sources of information available in London, found such a preparation very useful, and asked that their successors should also receive it, but such a makeshift training is obviously an inadequate solution of the problem and only staff with a working knowledge of South Africa's information position can be considered as meeting requirements.

The pattern of interlocking responsibilities that these paragraphs suggest is not a very spectacular one, but any progress along these lines should yield good dividends. And “dividends” is a significant word, for many difficulties could be overcome if there were a more general realisation of the necessity for spending appreciable amounts of money on publications, information apparatus and adequately trained information staff. The scientific world finds it fairly easy to swallow the camel of laboratory expenditure, but it still strains at the gnat of information-service expenditure.

Differences in International Arrangements for Financial Support of Information Services

N.F.GRELL

This paper[1] endeavours to show that there is no one formula for laying down the responsibilities of Government bodies, professional associations, universities, and research and industrial organizations to provide improved information services and to promote research in documentation. These responsibilities depend on the scientific standing of nations and individuals, on the geographical surroundings, on whether widely or little known languages are used, and on a number of other factors difficult to identify.

Equally, there is no one formula for laying down conditions constituting agreements for financial support to information and documentation services. They differ because of such varying factors as: *(a)* the level of organisation to be sponsored, *(b)* the scientific subject matter, *(c)* the financial capacity of the sponsor, *(d)* the legal status of the sponsored as well as of the sponsoring organisation.

Arrangements to develop new or support existing information and documentation services may therefore range from: *(a)* the unconditioned grant, to *(b)* semi-conditioned aid, or *(c)* conditioned aid, and to *(d)* technical assistance excluding financial support.

THE PROBLEM

Many new international organisations have been created since the end of the second World War. Their tasks and objectives are different but their methods very similar, in particular, those connected with the information and documentation responsibilities which they have to fulfil.

The United Nations and the U.N. Specialised Agencies such as UNESCO,

N.F.GRELL Technical Information Section, European Productivity Agency, Paris.

[1] The views and opinions expressed in this paper are those of the author only, and do not in any way commit the organisations consulted.

WHO, and ILO have undoubtedly given a great stimulus to documentation and information services all over the world, and regional organisations such as the Organisation for European Economic Co-operation and its branches, or the Communities of the "Six" in Europe, have had considerable effects which will now be enlarged by the creation of the Common Market and Euratom.

Apart from the information and documentation services which these organisations developed inside their own headquarters and which are run by their permanent staffs, outside bodies were used or even created to carry out special documentation tasks or to render continuing services. Considerable funds are thus being spent, and it is not always at all easy to find the forms and conditions of execution which will produce the highest dividends.

International organisations, as well as national bodies, are continually faced with this problem of how to ensure the best results when they go about subsidising documentation activities, or entering into agreements for the promotion of documentation, collecting of information, or whatever the particular objective may be.

Unfortunately very little has been published so far on this matter; on the contrary, there is much unnecessary secrecy about conditions of aids and forms of agreements. It seems all the more advisable, therefore, to open discussion on this matter, as this can only contribute to better solutions for the future.

APPROACH BY CASE STUDY METHOD

The author had the opportunity to study about twenty-five cases of agreements between supporting organisations (S.O.) and supported services, institutes or individuals (S.I.). Experience is thus obviously limited, both by the number of cases and by the fact that they cover only two international organisations.

All the cases studied relate to work carried out outside the supporting organisations on the basis of "agreements." Information and documentation services performed by staffs of international organisations have not been incorporated in this study. The case studies covered the origin and history of "projects,"[2] the elements of the agreements, results obtained and "comments."

For the sake of brevity, no full report on the individual studies will be given here, but rather a summary of the "lessons" obtained, together with suggestions for provisions to be foreseen when concluding agreements of this kind.

It is well understood by the author that the conclusion and formulation of

[2] The word "project" will henceforth be generally used for documentation and information activities carried out under contract.

contracts or agreements are legal questions, and of concern to lawyers. The latter, however, can render their beneficial services only upon a clear definition of what is wanted; they must therefore be given all the elements necessary. It is the aim of this paper to describe these elements on the basis of personal experience, for eventual use by other people working in similar positions.

Although the author essentially focusses his arguments on the provisions of the various agreements, he has endeavoured to include in his considerations other factors affecting results, both at the pre-contract stage, for example, the selection of partners, or after the signing of the agreement, such as inspection of progress of work, or final utilisation of reports.

PRINCIPAL PROVISIONS TO BE ESTABLISHED IN AGREEMENTS

When money is available for supporting information and documentation work or services, attention may be focussed on the provisions outlined below, which should be reflected in the agreements accordingly. Prior to the conclusion of the agreement, the selection of the S.I. is an intrinsic question; from the cases studied, one would be in favour of institutes or associations rather than private individuals generally. Excessive overheads charged by such institutes may, however, make other solutions preferable.

1. DEFINITION OF OBJECTIVE

This is a responsibility of the S.O., unless support is given in the nature of an unconditional grant or aid. Where the objective is a more general one, for example, improvement of bibliographical services, the more immediate goal, e.g., preparation of a manuscript on…, should be specified. In any case, it should be clearly stated whether the main objective is to produce a report, possibly for publication, or whether the report is considered to be a by-product only, as may be the case with seminars or training projects. This provision should help to prevent uncertainty about the results expected, and induce both partners to think ahead before concluding their agreement. In cases where it is difficult to estimate the scope of the project beforehand, it may be useful to split up the project into two parts with separate and independent agreements, the first project serving mainly to explore the total work involved.

2. DESCRIPTION OF WORK

This should be laid down by the S.O., after consultation and in agreement with the S.I., in considerable detail. It may in fact represent the planning of the work, methods of execution, provisional dates for parts of the work, provisions for sub-contracting, if applicable, together with form and contents of the final report, date of delivery and number of copies. This provision should help to avoid differences of opinion regarding the method of execution, scope of work, and possible achievements during the execution period.

3. GUIDANCE BY S.O. OR AN ADVISORY COMMITTEE

Responsible administrators of the S.O. should assist and supervise the execution of work continuously, particularly from the procedural side, so that they may detect delays or obvious time failures. Where they are not in a position to do this, the S.O. might appoint an advisory committee. This provision should help to avoid delays in the completion of work, a frequent complaint in documentation and information projects.

4. APPROVAL OF WORK PERFORMED

The S.O. normally retains the right to approve or disapprove of work performed, i.e., accepts or rejects manuscripts, reports, or services. In cases, however, where outside specialised bodies may be in a better position to judge the value of work carried out, it may be preferable to appoint, in agreement with the S.I., a body such as an arbiter, whose decision would be accepted as final by both the S.O. and the S.I.

Such a provision might be more practical than arbitration clauses relating to disagreements of another nature, such as financial disputes to be settled by the Court, a provision which is unlikely to be taken up and which might therefore be dispensed with.

Many of the kinds of agreements used are more elaborate in respect of protecting the S.O. against all kinds of moral or financial consequences, than in respect of guiding the S.I. towards the best ways of reaching the desired objective. Thus (1) to (4) above seem to deserve particular attention.

5. ROYALTIES OR OTHER WAYS OF PROFIT SHARING

One may wonder why S.I.'s are almost never offered some share of potential income. The inclusion of some sort of provision for profits might contribute to an interest on the part of S.I.'s to complete the work in time.

The provisions discussed below are, in fact, foreseen in most agreements, and they are listed here for the sake of completeness and for the expression of some observations.

6. CALCULATION OF FEES

The size of the sum to be paid to the S.I. has frequently given rise to very controversial opinions. In fact there is no solution applicable to every case. Some devices for calculating the amount corresponding to the work per formed are: estimate of manhours spent and paid for at rates comparable to those paid in the S.O. for a corresponding grade. With aids or unconditional grants: equivalent or percentage of the amount raised by the S.I. itself, or general appreciation of the goals attained by the S.I. In cases of non-conditioned aid or subventions, funds are often split up for the support of numerous small projects; it would seem preferable, however, to concentrate on fewer and bigger projects. Dependance of the amount to be paid on the length of the report or the number of days worked (without direct supervision by the S.O.) seems to be a doubtful method of calculation because it neither measures the real efforts invested nor the quality of the product. It seems to be most useful to obtain the advice of administrators who have been dealing with a great number of cases, watching their financial balance, and are therefore in a position to compare amounts spent in the past for individual projects.

When calculating fees, a breakdown between items is advisable, distinguishing between fees for experts and secretarial help, purchase of material, travel, and daily allowances, etc.

Where the agreement provides for sub-contracts, it is advisable to determine the repartition of the money to be spent.

7. CONDITIONS OF PAYMENT

In principle they should be attractive enough to appeal but cautious enough to prevent losses. Conditions will always vary according to the negotiations of individual cases. Normally not more than one third should be advanced, and possibly two thirds after approval of the work performed. Exceptions should be made only when close and continuous inspection of progress of work has been arranged.

8. OTHER PROVISIONS

Other provisions drawn from existing contract forms and which could usefully be included in agreements where appropriate are as follows:

(a) The S.I. shall bear any expenses occurring in the carrying out of its work.
(b) The S.O. acquires sole copyright in all countries and languages.
(c) The S.O. is entitled to publish the results of the project, in whole or in part, or to adapt the manuscript at its discretion.
(d) The S.I. is responsible for obtaining permission from owners of copyright on material which had been previously published, if S.I. desires to incorporate such material.
(e) The S.I. warrants that it has the right to grant the copyright of material and illustrations supplied by it under the agreement.
(f) The S.I.'s corrections to proofs in excess of 10% of the cost of setting will

be charged to it; proofs to be returned without undue delay.

(g) The S.O. shall not be bound to publish the manuscript.

(h) The S.I. is not to be considered as an agent or member of the staff of the S.O., is not entitled to any privileges, immunities, etc., nor authorised to commit the S.O. to any expenditure or obligation.

(i) The S.I. is responsible for any insurance which might be necessary.

CONCLUSION

1. The inclusion in agreements of all the provisos listed above, or others assembled from more sources, will not entirely preclude bad results; they may, however, contribute towards reducing the number of such cases.
2. Bearing in mind the provisions mentioned will certainly assist the planning of projects and concluding of agreements on which their execution is based.
3. The nature and form of agreements has a vital influence on the success or failure of information and documentation projects, and is not simply a medium to protect the interests of the S.O.'s. Precise provisions should therefore also be made relating to guidance during the execution of work.
4. It seems worthwhile to undertake a study of this kind on a larger scale, collecting information from more than two sources and extending the number of case studies.

Free discussion of the advantages and disadvantages of various contract forms for the support of information and documentation services, as well as the morale or policy behind them, should lead to a better identification of this problem, and might eventually result in a more enlightened attitude of governmental (and private) sponsors in respect of subsidising information and documentation activities.

Training for Activity in Scientific Documentation Work

GEORGE S.BONN

The purpose of this paper is (1) to present a record of world-wide facilities presently available for the training of interested persons for activity in scientific documentation work, (2) to try to identify the responsibilities of agencies and organizations to provide these facilities, and (3) to suggest additional or differently oriented facilities and responsibilities which may be required to meet current and future needs of the users of scientific information.

The scope of the survey is world-wide within the limits of available information from published reports or from correspondence with more space being given to those facilities which seem to be unique, outstanding, or newly reported, and with references being made to adequate published accounts of other good ones.

Facilities for training are taken to include facilities (1) for formal education in colleges and universities, (2) for regular instruction by competent professional organizations, (3) for practical training on-the-job or in workshops, (4) for home study by correspondence, and (5) for continuation learning through current publications and periodic conferences of professional societies. The training is to prepare persons for work as special librarians (or documentalists or information officers or whatever else they may be called) concerned in any way with scientific or technological information, including the production of services to utilize such information and the performance of research and development to improve and enrich it. Facilities for the necessary subject training in science or technology per se are not included, but it should be pointed out that such subject training is considered to be highly desirable in order to enjoy a successful career in the science-technology information field.

GEORGE S.BONN Graduate School of Library Service, Rutgers, The State University, New Brunswick, New Jersey.

INTERNATIONAL ORGANIZATIONS AND THEIR PROGRAMS

Before examining the various national training facilities around the world, it is proper to point out the very important contributions to science-information-work training being made by certain of the international organizations active in the area of documentation and librarianship.

Certainly the most enterprising on the most levels in the most fields in the most countries is the United Nations Educational, Scientific and Cultural Organization (Unesco) with headquarters at 19, avenue Kléber, Paris 16[e], France. Its present program includes *(a)* sponsoring meetings and conferences on a variety of pertinent and timely subjects in the field of scientific information (1); *(b)* sponsoring, or otherwise aiding in their establishment, courses and seminars on documentation, bibliography, and other basic topics of interest to scientific information specialists (2); *(c)* sponsoring scholarships and travel grants for study and research in documentation and library education (3); *(d)* helping to establish and to train personnel for documentation centers in various parts of the world (4); *(e)* publishing and distributing outstanding reports, periodicals, series, and other works so necessary and so useful as training aids and as information media on documentation and library activities, techniques, methods, and training throughout the world (5). In many instances the emphasis in the documentation work is on the scientific aspects of it, as, indeed, it usually is in most present-day discussion of documentation. Many, if not all, of these projects are carried out in collaboration with other international organizations or their national affiliates, or with other national or local groups (6).

Two other international bodies working in the area of scientific documentation are the International Federation for Documentation (FID), Willem Witsenplein 6, The Hague, Netherlands, and the International Federation of Library Associations (IFLA), c/o Bibliothèque Nationale, Berne, Switzerland. In 1948 these two federations set up a joint committee to look into the training and professional status of archivists, librarians, keepers of museums, and documentalists, with Mrs. Suzanne Briet of the Bibliothèque Nationale (Paris) as *rapporteur*. (Lumping these four groups together in this manner seems to emphasize their common collecting-and-keeping function rather than any collecting-for-use function more common perhaps in America and other places.) The final report presented by Mrs. Briet to the joint committee was published in April 1950 by Unesco (another example of its helpful collaboration) first in French (7) and then in English (8).

The report with its important annexes (chronological table and list of training establishments, model courses, and multilingual bibliography of recom mended handbooks and works) clearly demonstrates the problems, the rewards, the gaps, and the high lights in professional training as of the late 1940's, and presents the reasoned background for the recommendations offered for "action in the near future:" (1) a permanent council for professional training under the auspices of Unesco should be established, (2) exchanges of students and teachers should be intensified, (3) equivalences in cultural attainments should be developed to facilitate study abroad, (4) an international professional training review might be founded under Unesco, and (5) an international school for teachers and directors of training establishments should be considered. Certain of these suggestions have been put into effect, notably the one about exchanges of students and teachers and, to a limited extent, the one about developing equivalences of cultural attainments.

Mrs. Briet's report was presented in September 1951 at the Rome FID/IFLA conference along with a Rapport complémentaire bringing the results up to the end of 1950 (9). Other joint committee meetings were held in Copenhagen (October 1952) (10) and in Vienna (June 1953) (11); the 1954 meeting was to be in Zurich and was to include a week-long seminar for those who teach documentation, but after the Vienna meeting no further activities of the joint committee were reported according to both Mr. F.Donker Duyvis, Secretary-General of FID, and Mr. P.Bourgeois, President of IFLA.

Instead, as Mr. Duyvis puts it, "FID members considered the joint committee to be too big to be of any real value for the special needs for the training of documentalists. The FID Council therefor decided in 1953 to establish an FID committee: FID/TD 'Training of documentalists'" (12). The committee, apparently, had not been active enough to have made any reports until the September 1957 FID Council meeting in Paris where plans were announced to produce a loose-leaf booklet giving in some detail the programs of national courses for the training of documentalists (13).

In IFLA, on the other hand, "the opinion prevailed that really practical results would only be achieved if the training of librarians proper was treated separately from that of documentalists, archivists and museum staff," as Mr. Bourgeois explains it (14). So, at the 1956 Munich meeting of IFLA an *ad hoc* committee was formed to prepare a preliminary report on professional training for librarians to be presented at the September 1957 Paris conference. The report of this committee, prepared by Dr. Egger of the Swiss National Library, presents a very thought-provoking analysis of the interrelated problems of professional training and of recruitment of librarians, particularly of special librarians, and points out the usefulness and even the indispensability of international cooperation and exchange in solving the two-in-one problem. "Notre métier, qui est au service de la documentation," the report concludes, "se doit de faire le joint, par ce moyen, entre les techniques, les sciences et les cultures et, dans un futur voué au robot, il pourra merveilleusement contribuer, par sa tradition, à sauver, à préserver cette petite flamme qui brûle au fond de l'âme des hommes" (15).

Five questions were raised by the report as to the part that IFLA could take in improving the status, the professional aspects, the public acceptance, the training, and the exchange of librarians. Further study of these questions is being undertaken by a new IFLA committee under the chairmanship of Mr. Piquard, Director of the Libraries of the University of Paris and present president of the Association des Bibliothécaires Français; its report will be awaited with interest.

Both FID and IFLA, alone and in various combinations, sponsor meetings, conferences, seminars, and other informational programs in a number of countries on topics of timely or special interest to all types of librarians and documentalists (16), and each federation puts out many publication that are valuable as training aids and as information media about training, techniques, and new ideas in documentation and librarianship generally (17).

Two other international groups, the International Organization for Standardization (ISO) and the International Council on Archives (ICA), alone or in cooperation with Unesco, also publish materials of either informational or educational nature (18) which at least should be noted in this brief discussion of international bodies. One more specialized and, therefore, perhaps more pertinent organization has been formed so recently that it has not had much more time than to get a training committee appointed, but reports and suggestions should soon be coming from the International Association of Agricultural Librarians and Documentalists (19).

It may be noted in passing that the general feeling in these international groups seems to be that "although documentation may be considered a special branch of library service using special techniques, it should be as closely associated with other aspects of librarianship as possible," as Eileen R.Cunningham expressed it in her report on the Brussels 1955 International Congress of Libraries and Documentation Centers (21). This same observation can not be made as readily about many national groups, as will be evident in the following review of certain important ones of them.

TRAINING FACILITIES IN EUROPE

Since the comprehensive reports made by Mrs. Briet during the early 1950's, a few short surveys of education facilities for librarianship or documentation work have appeared (22), and there was the Carnovsky paper in 1948 which covered some of the history of library education in a number of countries (23). Then there is the Unesco series of annual reports on *Bibliographical services throughout the world* (24) which include a statement from each reporting country on the teaching of bibliography in that country, frequently specifically mentioning training facilities for science-technology bibliography, scientific documentation, special librarianship, and so on. All these cover more than one country and are often quite general, although they may contain the only available information on activities within some of the reporting countries, and may be, therefore, referred to on rare occasions in the following country-by-country survey of training facilities for scientific documentation work.

UNITED KINGDOM

In order to get into the library profession in the United Kingdom one must be certified as to professional qualifications by either (1) passing a series of rather stiff examinations and having the requisite experience and language ability (the usual way), or (2) graduating from the School of Librarianship and Archives in the University of London's University College (the unusual way). Professional standards and the examinations are set by the Library Association, the long-established (since 1877) national organization representing to some degree all library interests in the U.K. But it does not provide training facilities for those who wish to take its examinations. For these persons training is available (1) on the job at an approved library; (2) through personal study, attending lectures, and so on; (3) through correspondence courses conducted by the Association of Assistant Librarians, an affiliate of the Library Association; (4) at some 38 centers throughout the country which offer part-time courses; and (5) at nine post-war technical or similar-type colleges which offer full-time course programs (25).

Now, the organization directly concerned with special librarians, documentalists, and information officers is Aslib, and it is also very much interested in their professional training with the emphasis, naturally, on the special library or science information service aspect of the training. Accordingly, Aslib is working with the Library Association to produce an examination syllabus which will be accepted as suitable to *all* types of librarians, since the L.A. exams have come to be associated in people's minds with public librarianship only, thus affecting, understandably, the type of training available in the various courses and programs mentioned above. (As a matter of fact, the whole concept of education for librarianship of whatever kind seems to be under considerable scrutiny in the United Kingdom, and getting in—or keeping in—the information officers is just another part of the investigation (26).)

“Because of these difficulties,” writes Mr. Leslie Wilson, Aslib's Director, "Aslib has, for something like fifteen years, *(a)* been trying to get other institutions to provide full length training programmes for the people in its own membership, and *(b)* provided from its own resources short introductory and refresher courses for the numerous people who come into information work without full-scale librarianship training or qualifications" (27). In addition, Aslib branches and subject groups provide short courses from time to time on topics of local or timely interest (28).

There is some evidence that the efforts of Aslib and of a few of its more seriously concerned members who publish their views in strategic journals are having a gradual effect on the courses and programs mentioned earlier. At least two of the technical college library schools are now offering courses intended for information officers and they seem to be interested in offering still more. A direct report from the instructor of one of these courses (Mr. J. Farradane, a full-time scientific information officer himself, and responsible for an important abstracts journal which his company's research laboratory publishes) tells how it got started and what it might lead to:

> My course on abstracting at the North-Western Polytechnic is the first of its kind here, and was started experimentally on the suggestion of Mr. Sewell (of the School of Librarianship there) to see if I could implement the suggestions I had put out in various articles concerning the training required for information officers. The Polytechnic school is of course not tied to the Library Association requirements alone, and is interested to experiment…. The North-Western Polytechnic is as yet far from putting on an "information scientists" course [but] they have asked me whether I could personally take on another set of lectures, e.g., on patent literature (29). A little further on Mr. Farradane reports, "The Manchester school [College of Technology] *has s*tarted a course intended for information work," but he rejects the school's opinion, decided in debate at the first session, that there is "no difference between special librarians and information officers," an opinion also held, he reports, by the Library Association.

Besides the short courses, Aslib also organizes many meetings and conferences during the year, all having to do with some aspect of special librarianship and all providing individual members opportunities for further self-education and for exchanging ideas and views (30). In addition Aslib has a sizable publication program to keep its members (and others interested) well informed and up-to-date on new methods, techniques, training opportunities, and so on (31).

It must be pointed out that the Library Association, too, in this area of special librarianship, organizes meetings and conferences and puts out special publications through its Reference and Special Libraries Section, its Medical Section, and its University and Research Section (32).

NETHERLANDS

In the Netherlands, as in many other countries, there is one general library organization [Nederlandse Vereniging van Bibliothecarissen, NVB (Netherlands Association of Librarians)] and one special library or documentation work organization [Nederlands Instituut voor Documentatie en Registratuur, NIDER (Netherlands Institute for Documention and Filing)]. The NVB, again as elsewhere, also has a Section for Special Libraries (Sectie voor Speciale Bibliotheken, SSB) to look after the interests of special librarians and documentalists in the overall organization.

The two groups, NIDER and NVB, have set up a Central Examinations Board (Centrale Examencommissie van NIDER en NVB) which is, in effect, the responsible body in the Netherlands for professional standards and education in the area of science documentation and special librarianship. This board determines the requirements for examinations, sets the regulations for them, prepares them, administers them, and then to those who pass them it grants diplomas certifying the successful candidates as Special Librarians, Industrial Archivists, or Literature Researchers (Documentalists) (33).

Separate and distinct from this Centrale Examencommissie, NIDER and NVB/SSB have set up a Joint Committee on Training (Gemeenschappelijke Opleidingscommissie, GO) which is concerned with the actual training facilities for work in science documentation or special librarianship. The GO offers a regular scheduled program of lecture courses which are designed to prepare newcomers to the field for the certifying examinations of the Centrale Examencommissie, and it offers a group of correspondence courses for assistants and workers in various types of libraries to help them in classification, alphabetization, or title description, for example (34).

The teachers of the GO-organized courses, according to the GO catalog, are librarians of university and college libraries and of industrial and government libraries, archivists, workers in business and in other industrial organizations, and other experts in the fields which are taught. As Mr. Ir M.Verhoef, secretary of the GO, explains, “The courses are based on the principle of transfer of knowledge by those who are working in the field, to the new-comers” (35). Mr. Verhoef continues:

This has worked out to be very successful up to now, but as our courses are still growing, it is becoming more difficult to find enough experts who have sufficient time available to hold the lectures and correct the home work. Therefore we are now studying the possibility of founding an institute (library school) with professional teachers. But even then part of the courses will be given by non-professional ones.

The courses have a good reception in industry and governmental departments, nevertheless we are quite aware that the contents of the courses must follow closely the new developments in documentation. It is for this reason that for each lecture course a special committee is set up to advise the joint committee on alterations in the programs.

For additional comments on the need for a library school and for improvement in courses see remarks by Mr. Kessen and Mr. van Dijk in the August 1957 *Bibliotheekleven* (36).

A number of industries have in-service training programs for information officers (37) as well as for technical translators (38), and training for information officers is available at the Technische Hogeschool in Delft (39). Basic and introductory library courses, useful to special librarians and documentalists, are available at the Royal Library and the Public Library in The Hague and at the Public Library in Utrecht (40).

For additional information on training in the Netherlands for special librarians, information officers, and industrial archivists see articles by van Dijk (41), by van Dijk and Berkelaar (42), and by Dreese (43), all rather recent.

NIDER, like its counterparts in other countries, has an educationally useful program of meetings, conferences, and publications on timely and pertinent topics, and the entire program should be considered along with other training facilities (44).

DENMARK

By an act of May 25, 1956, the Danish Library School was officially established, to be “the State School for librarians of public libraries, and for librarians and library assistants of scientific and special libraries” (45). The act provides for the administration and the management of the school; it places the responsibility for the program of study, for regulations, and for entrance requirements; and it establishes the period of appointment (four years) of the necessary lecturers, at least one of whom “shall for a certain length of time have served at a scientific or special library,” a stipulation which must surely gladden the hearts of Danish special and scientific librarians.

An article on the new school in the July 1956 *Bogens verden* (46) states that the law makes a radical change in the training of librarians for scientific and learned libraries, and that special librarians will now have the chance of regular training. Heretofore, national science libraries took librarians with MA degrees and relied on short courses and lectures from the previous public librarian training program to impart the library knowledge required. The training for industrial librarians was always rather inadequate and somewhat dissociated from the training for other librarians, but on three occasions in the past few years they were able to take part in 74-hour basic courses in librarianship given by the Danish Association of Scientific and Special Librarians (47). Several continuation courses followed the basic courses, but presumably this series of courses will be discontinued once the Danish Library School gets into productive operation.

Education at the Danish Library School is free for all students, but the requirements for admission and for continued attendance are rigidly determined. Section II of the school is for the scientific and technical library training (Section I is for the public library training) and is open only to those who have permanent jobs in such libraries, either as librarians or as library assistants. The librarians get 250–300 hours of courses oriented toward administration and readers' services, while the assistants get about 200 hours of technical services courses. Special timely short courses and lecture series will be available occasionally for advanced study, and the school eventually will be expected to publish research papers in the area of librarianship generally (48).

SWEDEN, NORWAY

Perhaps the most active group in Sweden as far as special library training is concerned is the Society for Technical Documentation (Tekniska Litteratursällskapet, TLS) which has had a committee on professional training of industrial librarians since 1948. Besides arranging meetings and training courses, TLS has set up study groups and committees dealing with numerous technical library matters (49). Its extension courses, continuation courses, and workshops have been arranged from time to time for industrial librarians, industrial archivists, and for special librarians and documentalists from industry and research institutes, and a number of them have been described at some length in the Society's own internationally known journal, *Tidskrift för dokumentation* (50), and elsewhere (51).

Training for intermediate personnel of research and specialized libraries has been carried on since 1946 by the Royal Library in Stockholm under the auspices of the Association of Special Research Libraries, the Academy of Sciences, the Medical Library, and the Library of the Technological Institute. Practical and theoretical work is combined in a three or four months' course. For a good many years science libraries, university libraries, Technological Institute of Stockholm and of Gothenburg Libraries, the Royal Library, and the Archives have required only a university degree and a willingness to work through graded probationary periods of 3, 9, and 24 months in order to become a librarian on their staffs (52). The State Library School trains mostly for public librarianship, and the Stockholm Municipal Library organizes courses for augmenting its own staff, but special librarians and science documentalists may find useful basic courses either place.

Swedish librarians, like those in a good many other countries, have been trying for some time to achieve a uniform system of training which would apply to public librarians as well as to university and special librarians. A 1952 proposal suggested a four-year central library institute in Stockholm, and a 1956 proposal recommended that library science become a new subject department at the four major universities so that it may be more closely related to the academic field (53), but so far neither proposal has been acted on as far as can be determined.

The National Library School at Oslo, Norway, offers a nine months' course of training for positions in special libraries and in research libraries besides in the usual popular and children's libraries. Students must have university degrees and two years' experience in a recognized library. Certain large libraries have their own training programs, such as the three-year program at the University of Oslo library which is tied in with work toward a university diploma (54). There have been reports in the past few years that the National Library School was considering the possibility of giving courses only every other year, but there have been no recent announcements one way or another (55).

FRANCE

The Union française des organismes de Documentation (UFOD) is the agency in France concerned with training for science documentation work, special librarianship, or information work, and has been offering both preparatory and technical courses for “documentalistes” since 1944 (56).

The preparatory course (CPD) is for young assistants and leads to a certificate of proficiency. The technical course (CTD) takes two years to complete and is in two parts, the first consisting of instruction in general documentation, and the second in specialized documentation; it leads to an Institut National des Techniques de la Documentation (INTD) diploma and, after acceptable presentation of a thesis, to a State Diploma for Specialized Documentalists (57). Both the CPD and the CTD may be taken by correspondence, too. Descriptions of the courses and course syllabi are given in Annex III of the Briet joint committee report (58).

Elementary general library training courses have been organized by the Association des Bibliothécaires français; intermediate training has been available at the Library School of the Catholic Institute of Paris and through a supplementary course at the École des Chartes, the venerable school for archivists; and higher training has been organized under the authority of the ad ministration of the Bibliothèque Nationale (59). Certain of these courses will of course be helpful in the basic library training of special librarians or science documentalists.

Mr. R.Staveley, in his review of Malclès' *Cours de bibliographie,* took occasion to describe briefly France's new pattern of professional library training schemes of five important educational courses, in which special librarianship and information work are comprehended under the general term of Documentation (60).

CZECHOSLOVAKIA

The UFOD courses just mentioned, Mrs. Briet suggests, "may be said to have formed the basis of the training given in Czechoslovakia, Hungary, and Belgium" (61) for documentation work. The course in technical documentation given at the Polytechnic Library in Prague (62) does seem to follow the UFOD pattern, but there are other types of courses also available in Czechoslovakia.

Chairs of librarianship have been established since 1950 in the Faculties of Philology of both Charles University (Prague) and Comenius University (Bratislava) offering four-year non-diploma courses and five-year diploma courses as well as five-year correspondence courses. "For specialists employed in scientific libraries, the chair of librarianship arranges, in collaboration with the University Library, one-year courses covering a total of 200 to 240 lecture hours (instruction is given once a week for a full day) at which lectures in a condensed form are delivered on all the technical library subjects covered by the university course," according to Dr. Jaroslav Drtina, holder of the Chair of Librarianship at Charles University (63). In addition, continues Dr. Drtina, "In 1953 the chair of librarianship in Prague started a course for postgraduates at the highest level, open to holders of university degrees who have had two years' practical experience and intend to devote themselves to scientific work either at a university or in some other scientific sphere."

Short-term courses in librarianship and bibliography are available at the two large universities' libraries and at the libraries of the Academies of Science of Prague and of Bratislava (64). Still other courses, for the middle ranks of library workers, are offered at a number of the four-year vocational and technical schools throughout the country such as at Prague, Bratislava, and Brno (65). Among these more general courses there are a number which could be useful to special librarians.

HUNGARY

Since 1949 the National Library Center (Országos Könyvtar Kozpont), set up by a 1947 law, has organized professional documentation courses, presum ably patterned after UFOD's, which include library science, UDC standardization, cataloging, classification, abstracting, and service in information bureaus (66).

The *Unesco bulletin for libraries* in February 1956 reported:

In the 1948–49 academic year, training of librarians on a university level began at the Philology and Literary Science Faculty of the Lóránd Eötvös University at Budapest. At the same time the Library Science Institute was founded as a department of the university to meet the need for expert librarians in the large libraries and in scientific libraries (67).

A form of training had to be adopted that would satisfy the need for experts of scientific and public libraries alike, and the curriculum decided on includes basic library science courses, some other specialized science courses, and two modern languages in addition to Latin and Russian.

The National Technical Library (Országos Müszaki Könyvtar) of Hungary in Budapest has a very practical training device in operation through its Methods and Procedures Division. This division first works out suitable methods to ensure the satisfactory operation and development of the National library, and then it gives help on methods and procedures to other technical libraries and to specialized libraries in factories. For librarians assigned to factories, it organizes seminars on the special problems of this type of libraries, and it publishes an information bulletin for them under the title of *Information for the staffs of technical libraries* (68).

A new library law was passed in March 1956 which regulates the training and appointment of librarians and places all Hungarian libraries under the administrative control and charge of the Minister of Public Education (69).

BELGIUM

In August 1951 a Belgian Bibliographical Committee was established at the Ministry of Education, and it included besides members of the Bibliothèque Royale the secretary-general of the Belgian Documentation Association. Among other topics the committee has interested itself in professional education for librarianship and documentation and has recommended a plan to coordinate teaching at the four major library schools in the country: the Bibliothèque Royale (also responsible for examining and certifying librarians), the Ecole Provinciale de Bibliothécaires du Brabant, the Provinciale Bibliotheekschool at Brussels, and the École du Service Social also at Brussels. While these schools offer training for general library work, they also provide training for documentation work (especially the latter three) somewhat after the pattern of the French UFOD courses. There have been no reports of special courses for science documentalists as such, but basic library techniques training is avail able at these four schools and at others in Antwerp, Liége, Mons, and Hasselt (70), for science documentalists, special librarians, and others who may find them useful.

GERMANY

In the Federal Republic of Germany training for librarianship is centered more or less independently in Cologne, Hamburg, and Munich; in the German Democratic Republic training is systematically organized in the Berlin and Leipzig areas; but in neither one are there any special training facilities for documentalists or for librarians of scientific special libraries except on the job through staff training programs and through workshops and short courses arranged by professional societies. [See Mrs. Briet's reports for further concise information (71), Dr. Horst Kunze's handbook for a survey of general German library literature on all aspects of librarianship including professional training (72), a Bundesrepublik handbook for a list of the library schools in the western zone (73), and a Demokratische Republik volume for a description of the activities of its Library Commission for Professional Education (74).]

Much of the following information on German activities for training personnel in scientific documentation work comes from a special report in a letter from Prof. Dr. Horst Kunze, Hauptdirektor of the Deutsche Staatsbibliothek in Berlin (75).

In the Demokratische Republik the professional organization most concerned with training of documentalists has been the Zentralstelle für wissenschaftliche Literatur (Central Office for Scientific Literature) which was founded in 1951 but which has been since January 1, 1957, incorporated with the Institut für Dokumentation of the Deutschen Akademie der Wissenschaften zu Berlin. In 1953 the Zentralstelle formed an Arbeitsgruppe für Ausbildung und Arbeitsmittel which was to clarify the questions of basic training and the kind of training program, of the status of the profession, and of training materials and manuals. Documentalists, incidentally, were divided into three professional groups: scientific services, bibliographical services, and technical services. For more information on the work of this Arbeitsgruppe see the report by Mr. J.Koblitz in *Dokumentation* for 1953/54 (76).

The “General documentation” division of the Zentralstelle organized workshops for documentalists and bibliographers in a number of cities beginning with one in Freiberg/Saxony in December 1953, and continuing with others in Freiberg, Jena, Rostock, Dresden, Leipzig, Weimar, and Berlin; reports and announcements of these have appeared regularly in *Dokumentation,* as have method and summary reports of the General documentation division (77). Training of documentalists was also included in the workers' training centers of certain nationalized firms (78) and a 20-hour course on the decimal classification was given at a Berlin university by one of the members of the Zentralstelle. A series of books about practical documentation work, "Bücherei des Dokumentalisten," is being published by the Zentralstelle; of the eight titles so far, one, a manual by Peter Herrmann entitled *Practical application of the decimal classification,* is now in its third enlarged and revised edition and has found wide acceptance throughout Germany.

In the Bundesrepublik, training courses for documentalists have been organized by the Deutsche Gesellschaft für Dokumentation (German Society for Documentation) which announces and reports on the courses in its *Nachrichten für Dokumentation* (79); the Society was formed in 1941 and has committees on many aspects of documentation including training. In addition, Wilson reports that the Deutsche Normenausschuss (German Standardization Committee) holds periodic seminars for librarians on the development and application of the U.D.C. (80).

In special libraries throughout Germany the emphasis seems to be on staff training after the subject specialist is once on the job. This may include professional reading, practical experience for a period, or even temporary work in a larger scientific library. Two books have been published recently which should be of considerable assistance to administrators of special libraries with insufficient library training (81); and a number of articles have appeared on education, training, qualifications, and recruitment of special librarians and documentalists (82). Training in the fundamentals of library work for these persons, is of course available in the several centers mentioned at the beginning of this section on Germany.

POLAND

Special librarians in Poland are, to some extent, recruited from graduates of certain library schools which are attached to universities and have developed particular specialties: Warsaw University for research libraries and the University of Lodz for documentation, for example (83).

The Central Institute for Scientific-Technical Documentation (founded by decree in 1950) has overall responsibility in its area (84), and it organizes from time to time courses of lectures and exercises on the history, methods, and organization of bibliography and on documentation (85). But again, as in Germany, in-service training and guidance and encouragement from older librarians (86) are looked upon as necessary for the adequate professional training of special librarians (87).

There are, however, a number of other programs for general library training at public libraries in Warsaw, Cracow, Poznan, Lodz, and Wroclaw; at social service schools in Cracow and Poznan; at training centers in Jarocin and elsewhere (88); and through correspondence courses financed by the Ministry of Education (89), all of which could be useful to provide basic library training for entrance into science documentation work.

ITALY

Early last year Dr. Bruno Balbis, director of the Documentation Center of the National Research Council (in full, Centro nazionale di documentazione scientifica al Consiglio nazionale delle ricerche), began a course on the "Techniques of scientific information" in the Faculty of Statistics of the University of Rome (90). The course of 15 lectures covers such topics as the preparation of the scientific report, distribution of scientific publications, methods of dissemination of scientific information, the organization of the material, its documentation, mechanical and non-mechanical methods of retrieval, and related subjects (91). Visits and practical exercises go along with the lecture series.

A previous paper by Dr. Balbis on the development of documentation in Italy (93) offered suggestions for appropriate professional education for documentalists, but as yet documentation is not an official course of instruction in the library schools and programs of Italian universities (for example, Rome, Florence, Naples) which offer the usual courses for a general library education (94). But these could be helpful to those going into science documentation work, too.

SWITZERLAND

The following account of the types of training available in Switzerland to all librarians, including industrial and scientific librarians and documentalists, is from a letter written by Dr. P.Bourgeois of the Swiss National Library, Berne (95).

Two ways are open to candidates: the Library School in Geneva and the in-service training given in our libraries under the auspices of the Swiss Library Association. Both have the same program and are leading, after examinations and a diploma thesis, to equivalent diplomas.

The curriculum in the Library School is two years including at least two months practical work in a library. For the in-service training a minimum of 18 months' voluntariate is required before the examinations can be passed. Results are very satisfactory, and it is now quite exceptional that a library, even an industrial or administrative library, engages a new worker not possessing a diploma. For librarians in industry and administration, the Swiss Association for Documentation organizes from time to time special courses of two to four days, wherein particular subjects such as UDC, patent classification, punch cards, etc., are treated.

We have no school for candidates possessing a doctor's degree or another university degree. These aspirants to the library profession undergo the in-service training in a scientific library, a somewhat higher standard being asked from them at the examinations and for their diploma work than from the ordinary medium service candidates.

Wilson adds that the Swiss Association for Documentation also arranges the exchange of posts between documentalists for short periods (96), thus adding breadth and diversity to continuation learning programs.

UNION OF SOVIET SOCIALIST REPUBLICS

It has been reported in *Bibliotekaŕ* that up until at least 1953 there were no special training facilities in the U.S.S.R. to prepare workers for technical libraries on either a professional or a clerical level, there was no organizational or procedural guidance given to technical libraries by any agency competent to do so, and the large central technical libraries were backward in publishing instructional materials in the field (97). But by 1955 certain of the large scientific libraries (the State Scientific Library of the Ministry of Higher Education, the Gorkij Scientific Library of the Moscow Lomonosov State University, for example) were offering higher qualification courses for their own staffs as well as special bibliographical courses and seminars for scientific workers selected for technical library training, similar to those given by the Lenin State Library and the Saltykov-Schedrin State Library for the training of general librarians (98). The State Scientific Library also publishes textbooks and manuals to be used in its own, and other, training courses (99).

Library training facilities in the U.S.S.R. are of four types (100): (1) library *tekhnikums,* about 40, specialized vocational secondary schools which give a three-year general library training course; (2) library institutes, three, in Moscow, Leningrad, and Kharkov; independent four-year college-level schools which admit by competitive examination students who have completed their middle-school education, and which turn out specialist librarians and bibliographers of higher qualifications; (3) certain teacher colleges and the pedagogical faculties of certain universities, which may, like the library institutes, grant Ph.D.'s in library science (first one was given in 1944); and (4) libraries, which have one- or two-year programs of courses and seminars for general or specialized work. Basic training in any of these would of course be helpful to those who wish to work in scientific documentation.

In a brief letter (101) the director of the Institute of Scientific Information of the Academy of Sciences of the U.S.S.R., Prof. A.I.Mikhailov, indicates that all of their new employees must be specialists in natural science, in exact and technical science, and in several foreign languages. In the process of working these specialists are taught the necessary knowledge in the area of scientific information, documentation, and bibliography. In the near future the Institute was to have installed a special training course for scientific information work on a much higher level.

In July 1957 it was announced that a "Scientific research institute of librarianship and bibliography" was being organized in the Lenin State Library (102), but its specific duties were not spelled out nor was there any indication of how much if any training would be given there.

SPAIN

In Spain the preparation and training of librarians and documentalists is given by the *Cursos para la formación técnica de Archiveros, Bibliotecarios y Arqueólogos* held in the Biblioteca Nacional under the Direccion General de Archivos y Bibliotecas. Most students come to this professional training with a previous background of philosophy and letters, but candidates to become bibliographers or documentalists must have had both theoretical and practical training, they must pass a difficult test, they must know Latin and two contemporary languages, and they must have other educational requirements. In all, 32 regular courses are offered by the school (103) of which bibliographers and documentalists must take general reference bibliography, special bibliography, and documentation as three of the required six; they also must work for a time in the Información Bibliográfica y Documental section of the Biblioteca Nacional.

The principal problems in the area of scientific documentation in Spain, according to Mr. Francisco Esteve of the Biblioteca Nacional's Información Bibliográfica y Documental section, are *(a)* taking advantage of persons proceeding on scientific and technical careers, *(b)* employment of documental machines, and *(c)* possibility of using the same machines for bibliographical tasks (104).

In his letter Mr. Esteve also reports that there are other schools which teach more rudimentary or elementary courses in classification and documentation, one in Barcelona (Escuela de Bibliotecarias of the Biblioteca Central de la Diputación) and one in Madrid (Escuelas de Auxiliares de la Investigación of the Consejo Superior de Investigaciones Científicas). These could, of course, be helpful, too.

Elsewhere in Europe whatever training is given for scientific documentation work (or any other kind) seems to be available only in the libraries themselvse most concerned. Library associations frequently sponsor lectures and discussions on non-specialist problems, professional library journals publish occasional educational articles, and consultations and conversations among librari ans take care of many training problems, as in Yugoslavia, for instance (105). Dr. Duŝan Milaĉic, director of the Narodna Biblioteka in Belgrade, writes that staff members of scientific and industrial libraries in Yugoslavia are chosen from the best university graduates and are given in-service training as "dictated by the various functions the librarians have to do" through the actual working on the job as well as through attending meetings and lectures, reading journal articles, and so on (106).

TRAINING FACILITIES ELSEWHERE OUTSIDE NORTH AMERICA

ISRAEL

The new Graduate Library School of the Hebrew University, Jerusalem, opened in 1956, includes in its curriculum 42 hours on documentation, documentary reproduction, administrative problems of the special library, and preparation of bibliographies, according to a letter from Dr. I.Joel, its deputy director (107). More information on the activities of the new school is given in an article by Mrs. Nathalie Delougaz (108) who helped set up the school, and in a brief statement in the fourth annual *Bibliographical services* report (109). For background information on the school see Dr. Carnovsky's *Report on a programme of library education in Israel* (110).

Dr. Joel also reports that the Israel Institute of Technology (Technion) at Haifa is planning to offer training for special librarianship, but more precise information is not yet available. Heretofore, incidentally, suitable persons received only in-service training for (and by) working in Israel's scientific libraries (111).

UNION OF SOUTH AFRICA

Courses in documentation and/or bibliography and in special librarianship are available in the library schools of the Universities of Cape Town, Pretoria, Potchefstroom, and, by correspondence, South Africa (Pretoria) (112). The University of South Africa has recognized that because of the development of documentation and its association with scientific research it is necessary to have two types of training, (1) for research librarianship and (2) for public librarianship, with a portion common to each; details of its curricula are given in its journal *Mousaion* (113), and brief comments about the courses are given in other publications (114). Further recognition of the need for specialized training is seen in the fact that the South African Library Association, the national certifying body, includes special libraries and their organization in its final examinations (115).

The Library and Information Division of the South African Council for

Scientific and Industrial Research has held elementary seminars on technical library methods and on scientific and technical information in various centers (Cape Town, Durban, Port Elizabeth, Pretoria) and it has published a basic guide, *Books are tools* (116), on the organization of small technical libraries for South African industries (117). The former Chief Information Officer of the division, Miss Hazel Mews, has been since this past February Senior Lecturer in a newly organized Department of Librarianship in the University of the Witwatersrand, Johannesburg, so it is safe to assume that there are now additional facilities in the Union of South Africa for training in scientific documentation work.

AUSTRALIA, NEW ZEALAND

The New Zealand National Library Service sponsors the New Zealand Library School in Wellington, and there are correspondence courses also available, for general library training (118). Australia has several centers for general library training, in Melbourne, Sydney, Hobart, and Canberra among other places; and the Association of Special Libraries and Information Services (ASLIS) in Melbourne offers a number of training courses in various aspects of special librarianship and information service, and it also holds regular professional meetings of an educational and informational nature (119).

The State Library of Tasmania in Hobart has developed an in-service training program using its own senior staff (120), and the Commonwealth National Library in Canberra has a library school to train staff for itself and other government libraries (121). The Public Library of New South Wales in Sydney and the Public Library of Victoria in Melbourne also have their own library schools. These training curricula generally follow the pattern of the U.K. or of the U.S. so appropriate basic courses would be available to those planning to go into scientific documentation work, in addition, of course, to the more specialized ASLIS courses (122).

A fairly comprehensive article on “Australian libraries and their methods,” including, presumably, training methods, has been scheduled for publication in the January 1958 *Unesco bulletin for libraries;* it was prepared by Miss Barbara Johnston (123).

SOUTH AMERICA AND MEXICO

Argentina

In 1955 a library school was established in the Faculty of Philosophy and Literature of the University of Buenos Aires, an outgrowth of a school organized the year before by the Library Association of the Federal Capital which had been founded in December 1953. Bibliography and reference services, including science-technology, are courses incorporated in the curricula of the school's full program and short-course program. Late in 1955 the Centro Nacional de la Documentación Científica y Técnica was transferred back to the University of Buenos Aires under its old name of Instituto Bibliotecológico, the name it went under when it first instituted a course for documentalists from research laboratories, faculties, and state administrative services which, most likely, was continued (124).

Brazil

In March 1957 the Instituto brasileiro de Bibliografia e Documentaçâo began a series of courses planned especially to prepare specialized personnel to work in scientific libraries and information and documentation centers. Emphasis is on retrieval of scientific information, on research methods in the sciences, and on bibliography of the natural and medical sciences (125); basic library science subjects are also included in the training. Basic library subjects are taught at a number of places throughout the country, in library schools (São Paulo, Bahia, Campinas), in universities (São Paulo, Recife, Belo Horizonte, Curitiba), in the National Library in Rio de Janeiro, in the Department of Public Service in Porto Alegre, and in various state capitals by the Instituto Nacional do Livro (126).

Chile

The University of Chile gives a course in medical bibliography in its nine-month program for library education (127).

Colombia

A new Escuela de Bibliotecologia was opened in February 1957 at the Universidad de Antioquia, Medillín (128), offering at least the basic courses essential for library and documentation work.

Mexico

The Scientific and Technical Documentation Center organizes lectures for librarians on scientific documentation and lessons for scientists in French and in English (129), and the Escuela Nacional de Bibliotecarios y Archivistas devotes a considerable part of its bibliography teaching to the bibliography of science (130).

Peru

Courses in technical bibliography and in reference and research bibliography are given each year at the National School of Librarianship which is under the auspices of the National Library (131).

Venezuela

Bibliography of the sciences and of technology is covered in the three-year program of the School of Library Economy in the Faculty of the Humanities of the Central University (132).

Elsewhere in the world where there is any interest in scientific documentation, "the existing special librarians and documentalists have entered the field from conventional librarianship and have grown into their jobs, learning them by trial and error," as Mr. J.Saha, chief librarian of the Indian Statistical Institute puts it (133). Almost all the universities in India are running courses in librarianship, Mr. Saha reports, so basic library technique training is broadly available. One other country, Japan, should be mentioned because of its increasing interest in scientific documentation and the attendant increased work in that area. Some 40 universities in Japan offer lectures on basic librarianship and there are two college-level professional library schools which offer training for all types of library work. The Japan Library School in the Faculty of Literature of Keio-Gijuku University, and the Library Staff Training Institute of the Ministry of Education, both in Tokyo, include study of scientific-technological bibliography and reference materials in their curricula (134). The Japan Library School also does considerable consulting work with special libraries and sponsors workshops and clinics on various problems of interest to all kinds of librarians.

TRAINING FACILITIES IN NORTH AMERICA

CANADA AND THE UNITED STATES OF AMERICA

In the U.S. and in Canada much of the training for special librarianship or scientific documentation work is done in the 35 or so accredited college or university library schools in the two countries (two schools are in Canada) which meet the standards for a fifth-year college program of professional education for librarianship adopted by the American Library Association in July 1951. Graduation from any one of the accredited schools is comparable in effect to successful completion of the various registration requirements, examinations, and the like, for entrance into the library profession in the United Kingdom, the Netherlands, or elsewhere (135).

Actually there are 563 colleges and universities in the United States offering instruction of some sort in library science or bibliography, according to a survey prepared by the Office of Education of the U.S. Department of Health, Education, and Welfare (136), and there are at least another seven in Canada (137). Simple instruction in the general use of libraries and of reference materials is given in 233 schools; technical library training as a foundation for professional study, 8–12 hours of courses, is given in 278 schools; basic courses in library methods for part-time teacher-librarians, 6–23 hours of courses, are given in 220 schools; and an academic year of library study preparing for full-time positions, 24 or more hours of courses, is given in 122 schools, all these in the United States (138). Five of the seven in Canada offer an academic year of library study while the other two offer foundation training (139). In addition, six of the U.S. schools offer home study correspondence courses through their extension divisions (140), besides the courses mentioned above; in fact, a good many of the schools offer more than one kind of training as may be obvious from the fact that the numbers in the four groups add up to considerably more than 563, the total count.

A great many colleges and universities in these two countries also offer subject bibliography courses in science and technology (as well as in the humanities and in the social studies) given by the departments concerned; these courses are not included in the ones just enumerated, nor are any figures available on their numbers. But surely all the science-technology literature courses and probably most of the foundation library courses would be useful in some degree to persons going into scientific documentation work and to those already in, although, to be sure, non-accredited courses may not be acceptable to employers or applicable toward degrees.

In this part of the paper, however, only the facilities of the accredited library schools are being considered, and only those facilities that are particularly pertinent to the training of science-technology special librarians. Of the 37 accredited schools surveyed last year by the committee of which Dr. Karl Baer has been chairman (141), two are affiliated with state teachers colleges and are rather bound to be concerned only with school librarianship. A study of the catalogs and of the returns from the survey of the remaining 35 schools shows the following results:

(a) Eight schools have what may be called "special library training programs" mentioned or outlined in their published catalogs, including one in chemistry, one in science, one in science-technology, and two in medicine.

(b) Fifteen more have courses specifically identified with special or science-technology libraries, most of which seem to cover all types of special libraries.

(c) Three more schools have seminars arranged by kinds of libraries usually including special libraries in a separate seminar.

Of the 37 accredited schools surveyed in the spring of 1957, 26 had either a program of courses, an identified course, or a seminar in some aspect of special librarianship. All the schools have required bibliography courses, just as training programs in other countries have, which include some attention to the literature of the sciences and of technology often as part of a two- or three-course sequence of courses; thirteen schools have such an arrangement with a separate required course for science-technology literature as part of a sequence. Fourteen other schools have elective courses in science-technology literature.

Five schools give courses in "documentation" and two more in "science documentation." Two give courses in indexing and abstracting; and one in machine literature searching, language engineering, and applications of micro-recording. Five give courses in the bibliography of medicine; and one each in agriculture, biology, and pharmacy. One school has its own Center for Documentation and Communication Research.

All the schools [except four (142)] offer all or part of their curricula in summer sessions, at least five offer extension courses, and at least nine offer evening courses. Actual course offerings would naturally depend on interest, need, and demand. Furthermore, 26 of these schools offer courses, usually the basic or introductory ones, for undergraduate credit, making them more available to persons without the college degrees necessary to matriculate in a graduate program, although probably all the schools would arrange some "special student" category for those who wanted to take just certain courses. Several of the schools have "work-study" programs enabling students to work under guidance on a part-time schedule coordinated with the class schedule, thus combining on-the-job practical training with lectures and assignments in a graduate school, and getting paid for it.

For results of previous surveys of special library education facilities in library schools see Miss Gwendolyn Lloyd's 1954 report (143), Miss Linda Morley's 1947 report (144), and Dr. Jesse Shera's 1937 report (145).

There are several professional organizations, frequently including members from both Canada and the United States, which are very much concerned with science information work (or science-technology special librarianship) and with training for such work. Among these are the following especially pertinent groups:

(1) American Library Association (ALA) with its Committee on Accreditation (COA), Library Education Division (LED), and division-like Association of College and Research Libraries (ACRL) with *its* Pure and Applied Science Section (PASS) and Special Libraries Section (SLS); (2) Canadian Library Association (CLA) with its Research Libraries Section (RLS); (3) Association of American Library Schools (AALS); (4) Special Libraries Association (SLA) with its Science-Technology Division, Biological Sciences Division, Documentation Division, and Metals Division; (5) Medical Library Association

(MLA) with its Committee on Standards for Medical Librarianship with *its* Curriculum Subcommittee; (6) American Documentation Institute (ADI) with its Committee on Education; (7) Council of National Library Associations (CNLA) with its Joint Committee on Education for Librarianship, *its* Subcommittee on Special Library Education, and *its* Subcommittee on Course Content; (8) American Chemical Society (ACS) with its Division of Chemical Literature; (9) American Society for Engineering Education (ASEE) with its Engineering School Libraries (ESL) Committee. There probably are others of equal importance, either separates or affiliates, but these will give some idea of the variety (and of the complexity) of groups interested.

All these organizations carry on training programs for their members in the form of national and regional conferences and meetings and in the form of timely (145a) publications on new methods, new developments, new ideas, and so on. But none of these organizations sponsors or offers, as their European counterparts so generally do, regular courses or seminars to train persons for activity in science information work or for activity in any other type of work, although a few of them do hold occasional workshops or clinics or methods meetings on both national and local chapter levels. This fact should not be too surprising, though, when one considers (1) the American emphasis on school-sponsored and society-accredited professional education, (2) the American stress on strong national rather than regional organizations even though the important work is frequently done on the local level, and (3) of course, the very size of the area covered by Canada and the United States.

However, at least one of these groups, the Special Libraries Association, does have very active regional chapters, 31 of them including two in Canada, most of which have committees having to do in some way with training or education for special library work. Activity in this area at the same time varies greatly from chapter to chapter as is so evident from the results of a chapter survey made expressly for this paper in the fall of 1957 (146).

Seventeen chapters, including the two in Canada, reported in some detail what they have been doing in the area of education or training for special librarianship. Two distinct types of programs are in operation, one within the chapter for the benefit primarily of the members, the other outside the chapter for the benefit primarily of library school students or other non-members of the particular chapter.

Half the reporting chapters maintain liaison with nearby library schools: eight furnish speakers for talks on special libraries, eight (including three of the first eight) furnish the instructors for courses on special libraries (or at least the instructors are members of the local chapters), five arrange library tours for classes, nine invite interested students to meetings, and six act as consultants to the schools on special library course materials. Two chapters have been instrumental in establishing library courses in special schools, and one of these chapters has been instrumental even in establishing a special school with library courses. Two other chapters sponsor regional conferences jointly with other professional groups.

Training programs for members within the chapters are rather rare. Five of the 17 reporting chapters have never had such programs, but two of these are so newly organized that there has been no time for any. Six have had workshops or clinics at least once within the past two years (1956–1957). Four have had annual lectures by outside experts on educationally pertinent or timely subjects. Several have had at one time or another even fairly extensive education programs, but in most every case interest and participation could not be (or at least was not) sustained very long or renewed the following year, possibly because the particular series of lectures or discussions satisfied the immediate needs of those members who could benefit most from the program (147).

Most chapters have regularly published bulletins which frequently include items of immediate or local informational and educational interest. And one chapter has put out an experimental issue of a directory of training opportunities for special librarians in its own metropolitan area (148), certainly a long forward step in serving the training and educational needs of its members. Individual chapter concern for the training and education of its members is expressed in published reports (149) and in committee action and plans for programs (150), a few highly significant.

Even though these various formal and not-so-formal training facilities are available in the United States and in Canada, a number of science-technology libraries and science documentation centers find it expedient, or they may even prefer, to train their new employees on the job through in-service training programs of their own devising. Both professional and non-professional personnel, as judged by the type of work they do, are trained in such programs, and the programs naturally vary considerably in length, breadth, and depth depending on local needs.

For example, the Research Information Service of the John Crerar Library in Chicago employs a number of subject specialists to prepare comprehensive and exacting literature surveys, among other kinds of work they do. The determining factor in hiring one such person is his subject knowledge, given a general college education as background. Necessary library, bibliographic, or documentation techniques are learned on the job, by doing, under guidance. A comprehensive report by its director, Mr. Herman H.Henkle, covers the work of the Service and touches on its personnel problem (151).

Chemical Abstracts Service in Columbus has a similar problem, but rather in reverse, in that its problem is to train the indexers and editors of one of the abstracts journals which the John Crerar people use for their surveys. Again, subject knowledge (in this case, organic chemistry especially) is the specification most important to meet. On the job training in the *Chemical Abstracts* office is a five-year program of progressively more difficult and gradually less supervised work, beginning with relatively simple jobs very closely checked and ending eventually with indexing or editing any section of *Chemical Abstracts* on one's own. The problem, the training arrangements, and the special uses of equipment have been described at some length by Dr. E.J.Crane in 1955 (152).

Battelle Memorial Institute in Columbus with its 1800 or so scientists and research workers maintains a science reference library (or science information center) of some 35,000 volumes. The library staff of 70 professional and near-professional workers is part of a larger Information Management Division of 230 persons which in turn is part of the Economics Service of the Institute. All the workers in the library, except the clerical group, have science or technology college backgrounds, but only four of the 70 have library school training. They learn science information work on the job, as a member of a team. Battelle lays stress on teamwork.

A new man in the library's reference section will go through a prescribed schedule of jobs during the course of his in-service training which may take whatever time seems appropriate to the occasion, three years being not uncommon. The trainee is assigned for guidance to a library research group of seasoned personnel, and his introduction to the Institute is through the group leader. He tours, he reads, he works with different people. He makes preliminary searches, he handles telephone requests, he visits the branch libraries. He works for a time with other groups, abstracting, bibliography, cataloging, and ordering. Finally, then, he takes over his professional duties as reference information specialist in the Institute library (153).

There are many, many other (and simpler) approaches to in-service training in American science-technology libraries. A number of industries, for example, select technical persons from their staffs and send them to their main central libraries for training. Others hold quarterly or semi-annual conferences of all their technical librarians and information specialists to discuss basic and current problems and new methods. Some companies send out trained librarians from the main libraries to smaller libraries within the company to assist and to advise and to train the personnel (154). Most technical libraries will give training on the job in any specialized techniques required for particular positions (155). In any case, some instruction and indoctrination is always neces sary for new employees regardless of where else or how well they have been trained previously.

While these programs are designed for the training of persons who will work in particular libraries or information centers, such in-service training always stands them in good stead if they later work in other libraries. In the past few years a number of articles have appeared which discuss in-service training of professional staff (156), non-professional staff (157), recruits (158), or just staff (159).

PATTERNS OF TRAINING FACILITIES

Training for scientific documentation work, science information service, special librarianship, or whatever else it may be called, is available in varying degree throughout the world through three major kinds of training facilities: (1) library schools, (2) professional societies, and (3) in-service programs. Besides regularly scheduled facilities for training, these three also provide: (1) periodic meetings, conferences, seminars, discussions, clinics, workshops; and (2) current journals, manuals, other publications, for continuing education.

There is general agreement that, irrespective of the kind of training facility, the trainee should (1) have a science or technology subject background, (2) have or get suitable experience in a library preferably of the type he intends to work in, and (3) study fundamentals common to all types of libraries preferably along with other kinds of librarians, in addition to specialties useful in his later information or documentation work. But three difficulties come up, also generally agreed on: (1) how to attract trainees with such background, (2) what really are the fundamentals common to all types of libraries, and (3) how to get competent experienced instructors to teach the specialties and, as an afterthought, which specialties should be taught.

Professional library literature is full of partial, current, expedient, urgent, and otherwise variously described answers to these problems. But the literature also contains a number of decidedly thought-provoking books and journal articles from Germany (160), from the United Kingdom (161), from the United States (162), and from elsewhere (163), on these and allied problems.

Judging from study of available course outlines and taking into consideration varied meanings of descriptive course titles and terms, it can be stated that most of the training seems to cover pretty much the same type of material, at least on paper, with any variations in the methods, in the approaches, in the emphases, or in degree. There is, unfortunately, no easy way of finding out just what actually is being taught short of direct personal contact with each course and with each instructor; course syllabi are often not even set down on paper or available to outsiders if set down. A deeper study going beyond the paper descriptions now available would be profitable to identify trends and emphases and to pinpoint more precisely just what different schools, societies, or other groups (as well as individuals) mean by such terms as "documentation," "special librarianship," and "information work."

A parallel study of the work actually being done on the job by persons who call themselves special librarians, documentalists, information officers, information scientists, literature searchers, and whatever else, would also be profitable. Such a study would be most valuable, if not downright necessary, in order to determine the effectiveness of the presently available training for work in this broad area of science information activity and in order to suggest worthwhile changes in content or method for the improvement of present training. But beyond these findings, such a job analysis study along with the suggested course content study would provide working definitions of all these terms which are so glibly bandied about these days from the rostrums and in the literature concerned with science-technology information work. Since so much of the training seems to be so similar, it is quite possible that the jobs themselves are much nearer alike than some of their practitioners would have their readers and hearers believe. Neither job analysis nor course content could be covered in this paper.

PATTERNS OF RESPONSIBILITY

Local, regional, or national factors pretty well determine the form of training that will be available in any country. Certain of these, such as the legal basis for the existence of the profession, the development and influence of professional organizations in general, the system of education, tradition, and the acceptance and social status of librarians and documentalists, for instance, also will help determine where the responsibility, if any, lies for providing the training facilities that are or should be available.

Responsibility in a number of countries is established directly by law. The responsibility for all library activity or development may be assigned to some government agency; or a state school for library training may be organized, either separate from or attached to the national library. Responsibility is fixed, and needs and facilities can be determined on a national basis. Coordination is more possible and national standards may be established and maintained. Denmark, Japan, Hungary, Sweden, and the U.S.S.R. are examples of countries with library laws.

In several countries strong national library associations uphold high national standards by requiring certification through examination for entry into the profession. The United Kingdom, the Netherlands, and the Union of South Africa may be cited as examples. But the actual responsibility to provide the necessary training in order to pass the examinations has not been pinned down. Fortunately in every case the obligation to make training available to those who want it has been accepted by other professional agencies, library schools, and associations other than the ones which administer the exams. Since training is to help the student pass the examinations, the courses offered usually reflect the requirements of the examination syllabi. So the group of experts who prepare the syllabi determine, in effect, the content of the courses and therefore the direction of library education and development in these countries, inasmuch as no training would be needed or, presumably, given for those subjects or types of work not yet distinct enough (or respectable enough, or important enough, or whatever) to be included in the syllabi. Possibly thus indirectly the responsibility to provide facilities for training in scientific documentation work, for instance, lies with the exam and syllabi makers! It is certainly true, however, that the professional societies most concerned are offering courses to satisfy the needs of their members, and so are certain schools. But it is no wonder that there has been so much talking and writing about the Library Association syllabus and about library education in general in the United Kingdom.

In other countries, notably Canada and the United States, most of the facilities for training for a long time have been provided by library schools all of which in recent years have been attached to colleges or universities. Some 35 of these American schools, it was pointed out earlier, meet ALA's standards and are so accredited by the ALA Committee on Accreditation. All these accredited schools together make up the Association of American Library Schools. The Library Education Division of ALA is made up of ALA members who are interested in any way in formal education for librarianship, librarians, and library science teachers from *all* schools, not just the accredited ones. The AALS, then, is responsible for education for all types of librarianship in the U.S. and Canada through its member schools; and the ALA/COA is responsible for making sure the schools maintain high professional standards. But there is no guarantee that each accredited school *will* offer special courses for particular types of libraries which, in the minds of advocates of such special courses, might be considered necessary for adequate training for work in those libraries; on the other hand, each accredited school *may* offer such courses if it chooses, consistent with the ALA standards (164). The fact that so many library schools *do* provide specialized training in special librarianship and that largely in the area of scientific documentation work is an indication that there has been a demand from the profession, from students, from industry, and from within the schools themselves to make the courses available. The final responsibility to provide facilities for training in scientific documentation work, it then seems, rests with the people who are working in the field itself to make their needs and wants known, perhaps through appropriate professional society action. Since the LED "has specific responsibility for—1. Continuous study and review of changing needs for library education, development of educational programs, and continuing education of library personnel" (165), it may well be that the LED has some responsibility in this area, too; needs must be made known here, also.

In all countries, with or without other training facilities, a certain amount of in-service preparation is required of all library or information workers, ranging from brief indoctrination sessions to extensive well-organized courses. Every institution has not only the responsibility but also the right and the duty to train its own employees in the methods and procedures it expects them to use and in the philosophy and level of service it expects them to maintain; no previous training can possibly take the place of this kind of in-service training. Insofar as any institution is responsible for its employees' welfare, it is responsible for their having opportunities to improve their lots and to make themselves more useful and effective in their work; besides, it is profitable to the institution to have better trained and more efficient workers. So they have in-service training.

Again in all countries, with or without other training facilities, where there are well-established professional societies a certain amount of training is offered by the societies, ranging from workshops and methods meetings to full-fledged diploma courses for entrants into the profession. The societies usually hold meetings and conferences for all the members at least once a year at the national level and rather more frequently at the regional or local chapter level, for the exchange of opinions and ideas, for discussion, for fellowship, for enlightenment, and for transacting the business of the associations. The societies publish journals of informational and educational interest to keep their members informed on new developments, new concepts, new techniques in the profession, and handbooks and manuals to aid, guide, and assist members and others in the operation and administration of their libraries or information centers. The societies maintain headquarters staffs to serve the more immediate needs of the members (answering questions on training facilities, for instance) and to administer the activities of the association, among other jobs. The societies, in short, provide the clearing houses of professional knowledge, their *raison d'être*. In a recent article on "The functions of the professional association," Mr. Robert K.Merton has said that the professional association "is typically dedicated to the objective of raising the standards of professional education, for if the organized body of informed professionals does not take on this task, who else can?" (166) "The foremost obligation of the association," he says later on, "is to set rigorous standards for the profession and to help enforce them; standards for the quality of personnel to be recruited into the profession; standards for the training and education of the recruit; standards for professional practice; and standards for research designed to enlarge the knowledge on which the work of the profession rests." (167) In carrying out part of this obligation, Mr. Merton suggests, the association works to help prepare the practitioner for the more effective discharge of his professional roles and, acting on the philosophy that professional education is a life-long process, it establishes institutes to advance the education of the practitioner and helps motivate him to develop his skills and to extend his knowledge. Certainly most, if not all, professional associations in the area of scientific information work are aware of and are discharging their professional responsibilities to a remarkable degree.

In the last analysis, of course, the responsibility to provide training for any profession rests with each member of that profession, acting alone or through duly organized associations, joint committees, advisory bodies, or whatever. How well each member meets this responsibility is a measure of his concern for and of his devotion to his chosen profession.

DESIDERATA

One of the purposes of this survey of facilities and responsibilities to provide training for scientific documentation work was to locate both strong spots and weak spots among them with a view toward overall development and improvement. Most of the following suggestions for consideration were prompted by reviewing those strong and weak spots: to promote and encourage and spread the strong ones, and to remove and correct and discourage the causes of the weak ones. A few suggestions came from correspondents and other professional persons with whom the problem was discussed during visits to libraries, library schools, library associations, and librarians, all in the area of scientific documentation work or special librarianship.

INTERNATIONAL ORGANIZATIONS

1. Improve the interchange of ideas and opinions on education and training for work in the scientific information field through already existing and/or entirely new journals, manuals, handbooks, reading lists, and the like.
2. Work to improve the "climate" of librarianship generally so that much-

needed science-technology-trained workers will be more readily attracted to the field:

a. Improve social acceptance of library work of whatever kind.
b. Improve research, scholarly, and educational acceptance of libraries.
c. Improve salary scales.
d. Improve professional standards.
e. Improve professional status vis-à-vis other professional groups.

3. Encourage the interchange of teachers, of students, of working librarians in the area of scientific documentation work, special librarianship, and so on.
4. Encourage the establishment of an international school to train teachers of subjects and techniques involved in scientific documentation work.
5. Continue the present good work.

NATIONAL ORGANIZATIONS

1. Improve relationships and liaison generally on all levels between the association and its members and the existing schools and courses in the area of scientific documentation work and special librarianship in order to work together toward improvement and extension of training facilities; offer help, suggestions, instructors, materials, tours, student memberships; encourage and patronize extension, evening, summer courses; plan joint projects; exchange speakers and lecturers.
2. Encourage each chapter or local affiliate group to find out and publish information on nearby courses or other facilities for training in the general area of scientific documentation work or special librarianship, for the obvious benefit of its own members and of anybody else who might be attracted to the field.
3. Arrange short courses, workshops, seminars, panel discussions, exhibits, lecture series on timely topics, taking advantage of the interests and abilities of members, of nearby school facilities, of nearby governmental facilities, of commercial facilities (e.g., IBM, Rem-Rand); sponsor lecture tours of experts in the scientific documentation and special library field, as the ACS does.
4. Encourage joint meetings and programs with other professional groups in areas of mutual interest in order to broaden contacts and information both ways; encourage unity and work together on similarities within the whole profession.
5. Do what can be done to help with points 1 to 5 under international organizations.

LIBRARY SCHOOLS, SCHOOLS OFFERING LIBRARY COURSES

1. Improve relationships and liaison generally between the school and its instructors and nearby special librarians, information officers, and scientific docu

mentalists; encourage interest in courses, ask for suggestions, invite speakers, offer consultation; schedule extension, evening, summer courses; arrange institutes and short courses; offer high-level courses in areas of interest to science information specialists and special librarians generally; join local chapters.

2. Publicize more widely and effectively what the school has to offer in the way of courses that may be of interest to special librarians, science documentalists, and so on, both inside *and* outside the library science department; besides the library science courses might be mentioned courses in language, business, economics, subject literature, general science survey, research methods, basic science or technology, etc.
3. Encourage research and thesis topics on problems peculiar to special librarianship or scientific documentation work, including the literature of the field, the use of the literature, storage and retrieval of information, manual and mechanical methods of literature searching, the use of technical libraries, handling of questions in technical libraries, research methods in technical libraries, and education for work in the field. [See Dr. Herbert Coblans' paper in the July 1957 *Unesco bulletin for libraries,* and Miss Barbara Kyle's paper in the August 1957 *Review of documentation,* for some specific problem areas that need working on (168)].
4. Make it possible for *all* students in the school *(a)* to learn that work in special libraries, scientific documentation, and information centers does exist and is highly rewarding in many ways; *(b)* to learn in what ways and to what extent such work differs from other types of library work; *(c)* to visit at least one good example of a special library or science information center; *(d)* to elect a course that pays more than just passing attention to special libraries or scientific documentation work, a course in science literature, or in documentation, or in special library service, or perhaps a seminar in some part of the field, but at least some course in which problems peculiar to the field are taken up and discussed at some length.
5. Do what can be done to help with points 1 to 5 under international organizations.

LIBRARIES, INDUSTRIES, EMPLOYERS OF SPECIAL LIBRARIANS AND DOCUMENTALISTS

1. Encourage workers to take advantage of nearby education and training facilities to develop and improve skills, to learn about new techniques, to raise their educational level, to increase their chances for advancement, etc.
2. Develop in-service programs if the staff is large enough to encourage the advancement and participation of all the members of the group in all phases of

work. But even small staffs need some indoctrination besides just reading staff manuals.

3. Allow time for professional development, study, reading, courses, and the like; encourage participation in professional associations and at their meetings.
4. Encourage closer liaison between science information workers and other research workers for mutual development within the overall organization.
5. Do what can be done to help with points 1 to 5 under international organizations.

SPECIAL LIBRARIANS, SCIENCE DOCUMENTALISTS, INFORMATION OFFICERS

1. Join and participate in professional associations, library, special library, documentation, information, scientific, technical, or whatever.
2. Investigate nearby training facilities, educational, commercial, or governmental, for possible mutual or personal help and improvement; keep up with developments in both the library and the subject fields through current reading and study.
3. Contribute ideas and experiences to professional publications, to make them more educationally useful.
4. Think about, talk about, write about, and encourage improvements in any and all existing training facilities or methods or techniques or concepts or materials in the area of special librarianship or science documentation work or allied fields.
5. Do what can be done to help with points 1 to 5 under international organizations.

REFERENCES

1. CAIN, JULIEN. The development of bibliographical and documentation services over the past ten years. *Unesco Bulletin for Libraries,* v.10.11–12, Nov.–Dec. 1956. pp. 262–270. Chronology, pp. 269–270, covers conferences and meetings. International exchange of publications in Latin America, Meeting of experts …. *Unesco Bulletin for Libraries,* v.11.2–3, Feb.–Mar. 1957, pp. 29–31. Conference on classification for information retrieval, Dorking, England…. *Unesco Bulletin for Libraries,* v.11.4, April 1957, p. 97.
2. CAIN. *op. cit.,* p. 270. Seminar for documentalists, at Insdoc, New Delhi, India, May 31, 1956. *Unesco Bulletin for Libraries,* v.11.2–3, Feb.–Mar. 1957, p. 68. Seminar for documentalists, at the Technical and scientific documentation centre in Cairo, Egypt, June 1956. *ibid.*
3. Unesco's library field missions and fellowships. *Unesco Bulletin for Libraries,* v.10.11–12, Nov.–Dec. 1956. pp. 279–285. Fellowship holders on pp. 284–285. Library bursaries for research in the Council for Scientific and Industrial Re

search Laboratories, South Africa. *Unesco Bulletin for Libraries,* v.11.4, April 1957, p. 98.

4. Unesco's library field missions and fellowships, *op. cit*. Central American documentation centre for industry, Guatemala. *Unesco Bulletin for Libraries,* v.11.8–9, Aug.–Sept. 1957. p. 215. Scientific documentation centre in Pakistan, *ibid.,* p. 216.
5. The bibliographical and reference publications of Unesco. *Unesco Bulletin for Libraries,* v.10.11–12, Nov.–Dec. 1956. pp. 290–294. *Unesco bulletin for libraries,* 1947–date. Also in French and Spanish. CAIN. *op. cit.,* pp. 268–269. MALCLES, L.N. *Bibliographical services throughout the world,* First and second annual reports, 1951–52, 1952–53. Unesco, Paris, 1955. COLLISON, R.L. *Bibliographical services throughout the world,* Third annual report, 1953/1954. Unesco, Paris, 1955. ——, Fourth annual report, 1954/1955. Unesco, Paris, 1957.
6. CAIN. *op. cit.,* p. 266. Unesco's library field missions and fellowships, *op. cit*.
7. BRIET, SUZANNE. *Enquete sur la formation professionnelle des bibliothécaires et des documentalistes.* Unesco, Paris, 1950.
8. ——*Enquiry concerning the professional education of librarians and documentalists.* Unesco, Paris, 1951.
9. ——La formation professionnelle des bibliothécaires et des documentalistes, rapport complémentaire. *Revue de la documentation,* v.18.1, 1951. pp. 21–27.
10. ——L'Enseignement et le statut professionels des bibliothécaires et des documentalistes. *Revue de la documentation,* v.19.2, 1952. pp. 21–26.
11. ——Rapport à la commission mixte de l'enseignement et du statut professionnels…. *Revue de la documentation,* v.20.2,1953. pp. 49–56. See also v.20.4, 1953, pp. 132 and 135–136.
12. DUYVIS, F.DONKER. letter dated December 2, 1957, from La Haye.
13. *Ibid.*
14. BOURGEOIS, P. letter dated November 28, 1957, from Berne.
15. EGGER, E. Formation professionnelle et problèmes actuels du recrutement des bibliothécaires. Typescript of report, to be published in *Libri*.
16. CAIN. *op. cit.,* p. 266. WILSON, LESLIE. Organizations in the special library field. Chapt. 13 in *Handbook of Special Librarianship and Information Work,* Wilfred Ashworth, general editor. Aslib, London, 1955. pp. 356–359.
17. WILSON, *ibid*.
18. WILSON, *ibid.,* pp. 354, 360. MALCLES. *op. cit.,* pp. 297–299.
19. COLLISON. Fourth annual report, *op. cit.,* pp. 133–134.
20. Deleted on revision.
21. CUNNINGHAM, EILEEN R. International congress of libraries and documentation centers. *Special Libraries,* v.47.1, Jan. 1956. pp. 32–34.
22. e.g., PIETSCH, E. Nationale dokumentationseinrichtungen. *Nachrichten für dokumentation,* v.6.1, March 1955. pp. 5–18. See also other reports by Mrs. Briet, in *Bulletin de l'union française des organismes*

de documentation, v.22.5, Sept.–Oct. 1954, pp. 1–9, and in the *Reports* of the 1955 International Congress of Libraries and Documentation Centers.

23. CARNOVSKY, LEON. Education for librarianship abroad, in *Education for Librarianship,* edited by Bernard Berelson. ALA, Chicago, 1949. pp. 66–89.
24. MALCLES. *op. cit.* COLLISON. *op. cit.*
25. WILSON, *op. cit.,* pp. 334–337. Library association *Year book* and *Students' Handbook* give particulars in their annual editions.
26. e.g., PATON, WILLIAM B. Education for librarianship. *Library Association Record,* v.58.12, Dec. 1956. pp. 460–464. BENGE, R.C. Reactionary thoughts on the education of librarians. *Library Association Record,* v.59.1, Jan. 1957. pp. 2–4. MALLABER, KENNETH A. Education for librarianship. *Library Association Record,* v.59.3, March 1957. pp. 91–94. “A rejoinder to Mr. Paton” above. MORRISH, P.S. Training of librarians. *Library Association Record,* v.59.3, March 1957. pp. 104–5. Correspondence about Mr. Benge's article above. *etc., etc., etc*.
27. WILSON, LESLIE, letter dated October 15, 1957, from London. List of proposed short training courses also enclosed; see Appendix 1. Announcements of the courses are made in the “Aslib information” section of *Aslib proceedings* at regular and appropriate intervals, e.g., v.9.9, Sept. 1957, pp. 291–2.
28. e.g., Aeronautical group, Textile group, announced in *Aslib proceedings,* v.9.11, Nov. 1957. p. 352.
29. FARRADANE, J. letter dated October 16, 1957, from Kent.
30. WILSON, letter *op. cit.*
31. WILSON, in Ashworth, ed., *Handbook, op. cit.,* pp. 329–334. (ref. 16)
32. *Ibid.,* p. 336.
33. CENTRALE EXAMENCOMMISSIE (NIDER, NVB). *Examenreglement en exameneisen voor speciale bibliothecaris, bedrijfsarchivaris, literatuuronderzoeker.* The Commissie, 's-Gravenhage, 1955.
34. GEMEENSCHAPPELIJKE OPLEIDINGSCOMMISSIE (NIDER, NVB). *Cursussen 1957–1958;* opleiding van bibliotheek, archief, documentatie personeel. The Commissie, 's-Gravenhage, 1957.
35. VERHOEF, IR.M. letter dated October 21, 1957, from The Hague.
36. KESSEN, A., VAN DIJK, C., and others. Symposion over de opleiding. *Bibliotheekleven,* v.42.8, Aug. 1957. pp. 213–226.
37. PIETSCH, E., HANTKE, G., MULERT, G. Massgeblichte dokumentationseinrichtungen in den Niederland. *Nachrichten für Dokumentation,* v.5.1, March 1954, pp. 13–24.
38. CITROEN, I.J. Training technical translators. *Babel,* v.1.2, Dec. 1955. pp. 61–64.
39. PIETSCH, HANTKE, MULERT. *op. cit.*
40. BRIET. *Enquiry….* *op. cit.,* p. 28. MALCLES. *op. cit.,* p. 220.
41. VAN DIJK, C. Special library training as it is in the Netherlands today. *Aslib proceedings,* v.5.4, Nov. 1953. pp. 276–283.
42. VAN DIJK, C. and BERKELAAR, P.H. The training of information officers in the Netherlands. International Congress of Libraries and Documentation Centers, 1955, *Reports,* II B, pp. 90–93.

43. DREESE, M.J. De opleiding tot bedrijfsarchivaris. *Tijdschrift voor efficientie en documentatie,* v.26.4, 1956. pp. 123–125.
44. WILSON, in Ashworth, ed., *Handbook*…. *op. cit.,* p. 350.
45. Copy of translation of the law, from Dr. Preben Kirkegaard, Director of the Danish Library School.
46. Danmarks biblioteksskole. *Bogens verden,* v.38.4, July 1956. pp. 181–185.
47. HØRDING-JENSEN, TOVE. Firmabibliotekarer og deres uddannelse. *Tidskrift för dokumentation,* v.11.3, 1955. pp. 31–33.
48. Copy of translation of a report on the school, from Dr. Preben Kirkegaard, Director of the Danish Library School.
49. WILSON, in Ashworth, ed., *Handbook*…. *op. cit.,* p. 353.
50. FAHLSTRÖM, JAN MAGNUS. Course for industrial archivists (in Swedish). *Tidskrift för dokumentation,* v.10.4, 1954. p. 41. Stimulerande TLS kurs. *Tidskrift för dokumentation,* v.12.3, 1956. pp. 42–44. ENGEL, ERWIN. Bibliografiundervisning för företagsbibliotekarier. *Tidskrift för dokumentation,* v.12.5, 1956. pp. 59–61.
51. OTTERVIK, G. and MOHLENBROCK, S. Training for library service in Sweden. *Annals of Library Science,* v.1.1, March 1954. pp. 59–64.
——and ANDERSON, I. *Libraries and Archives in Sweden.* Stockholm, Swedish Institute, 1954. includes information on training for library service.
52. BRIET. *Enquiry*…. *op. cit.,* p. 30.
53. Utbildningsfragan. *Biblioteksbladet,* v.41.8, 1–2, 1956. pp. 589–621.
54. GERDWAN, Mrs. Library education in Norway. *Abgila,* v.3.4, Dec. 1953. pp. 175–178. BRIET. *Enquiry*…. *op. cit.,* p. 27.
55. COLLISON. Third…. *op. cit.,* p. 62; Fourth…. *op. cit.,* p. 81.
56. MALCLES. *op. cit.,* pp. 65–66.
57. *Ibid.,* p. 66.
58. BRIET. *Enquiry*…. *op. cit.* Annex III, pp. 28–35.
59. BRIET. *Enquiry*…. *op. cit.,* pp. 35–36.
60. STAVELEY, R. in his review of *Cours de bibliographie*…. *Journal of Documentation,* v.11.2, June 1955. pp. 83–84.
61. BRIET. *Enquiry*…. *op. cit.,* p. 43.
62. *Ibid.,* p. 38.
63. DRTINA, JAROSLAV. Library training in Czechoslovakia. *Unesco Bulletin for Libraries,* v.11.4, April 1957. pp. 80–82.
64. COLLISON. Fourth…. *op. cit.,* p. 62.
65. SOLCOVA, SVATAVA. The Czechoslovak library system. *Unesco Bulletin for Libraries,* v.10.7, July 1956. pp. 151–154.
66. BRIET. *Enquiry*…. *op. cit.,* p. 25.
67. Library development in Hungary. *Unesco Bulletin for Libraries,* v.10.2, Feb. 1956. pp. 34–35.
68. National technical library of Hungary. *Unesco Bulletin for Libraries,* v.10.7, July 1956. pp. 159–160.
69. KOVACS, MATE. The aims, basic principles and cultural significance of the Hungarian library law (in Hungarian). *Magyar könyvszemle,* v.72.3, July– Sept. 1956, pp. 181–200.

70. BRIET. *Enquiry…. op. cit.,* p. 22. MALCLES. *op. cit.,* p. 54.
71. BRIET. *Enquiry…. op. cit.,* p. 19.
72. KUNZE, HORST. *Bibliotheksverwaltungslehre.* Leipzig, Harrassowitz, 1956.
73. VEREIN DEUTSCHER VOLKSBIBLIOTHEKARE. *Handbuch der offentlichen Bücherein.* Hamburg, Eberhard Stichnote, 1952.
74. BIBLIOTHEKSKOMMISSION FÜR AUSBILDUNGSFRAGEN BEIM STAATSSEKRETARIAT FÜR HOCHSCHULWESEN. *Aus der Arbeit der wissenschaftlichen Bibliotheken in der Deutschen Demokratischen Republik.* Leipzig, Harrassowitz, 1955.
75. KUNZE, HORST. letter dated January 7, 1958, from Berlin.
76. KOBLITZ, J. *Dokumentation,* v.1.4, 1953/54. pp. 91–92, 123–124. (from Kunze).
77. *Documentation,* v.1.5, 1953/54, pp. 112–113; v.1.6, 1953/54, p. 127; v.2.1, 1955. p. 16; v.2.2, 1955, p. 39; v.2.3, 1955, p. 58; v.2.5, 1955, p. 101; v.3.3, 1956. p. 64–65; v.3.6,1956, p. 137–138 for announcements and reports of workshops, and v.2.3,1955, p. 58; v.4.2,1957, p. 31–37 for summaries. (from Kunze)
78. *Gesetzblatt der DDR* 1953, v.1, p. 406. *Arbeits- und Sozialfürsorge,* v.8, 1953, p. 581. (from Kunze)
79. *Nachrichten für Dokumentation, passim,* e.g., v.7.2, June 1956, p. 100–101.
80. WILSON, in Ashworth, ed., *Handbook…. op. cit.,* pp. 352–353.
81. BRÄMER, JOACHIM, and VOGEL, DIETER. *Die wissenschaftliche Fachbibliothek….* Leipzig, Harrassowitz, 1956. FUCHS, HERMANN. *Kurzgefasste Verwaltungslehre für Institutsbibliotheken.* Wiesbaden, 1957. (from Kunze)
82. GÜLICH, WILHELM. Ausbildungs- und Nachwuchsfragen in Bibliotheken und Dokumentationsstellen. in *Arbeitsgemeinschaft der technisch-wissenschaftlicher bibliotheken, Bericht über die 5-tagung in Braunschweig am 26 und 27 März 1953….* Essen, 1953. pp. 61–76. GRUNWALD, WILHELM. Technische Hochschul-Bibliotheken. *Zeitschrift für Bibliothekswesen und Bibliographie,* v.2.4, 1955. pp. 257–279. see part 3, esp. STÖTZER, WALTHER. Industrielbibliotheken und Industrielbibliothekare. *Zeitschrift für Bibliothekswesen und Bibliographie,* v.2.4, 1955. pp. 279–292. FABIAN, RUDOLPH. Ausbildung von Nachwuchskräften in industriellen Dokumentationsstellen. *Dokumentation,* v.2.6, Nov. 1955, pp. 117–119.
83. DEMBOWSKA, M. Prerequisites of professional growth of younger staff members (in Polish). *Bibliotekarz,* v.33, Feb. 1956. pp. 26–28. BRIET. *Enquiry…. op. cit.,* p. 29.
84. SIEROTWINSKI, S., tr. O.Wegner. Die wissenschaftliche-technische Dokumentation in der Volksrepublik Polen. *Dokumentation,* v.1.8, Sept. 1954. pp. 154–159.
85. COLLISON. Third…. *op. cit.,* pp. 65.
86. DEMBOWSKA. *op. cit.*
87. PRZELASKOWSKI, R. Library staff in research libraries (in Polish). *Przeglad biblioteczny,* v.24, Jan.–Mar. 1956. p. 24.36.
88. BRIET. *Enquiry…. op. cit.,* pp. 29.
89. LACOUR, CHRISTINA. “Remarques sur le service des bibliothèques dans le Republique Populaire de Pologne.” *Archives, bibliothèques, collection, documentation,* no. 11, Sept. 1953, pp. 296–298.

90. Instruction in scientific documentation at the University of Rome. *American Documentation,* v.8.2, April 1957. p. 146. BALBIS, BRUNO. letter dated January 9, 1958, from Rome.
91. Programma del corso di Tecnica dell 'informazione scientifica svolto dal dott. Bruno Balbis, direttore del Centro nazionale di documentazione scientifica al Consiglio nazionale delle ricerche, presso la Scuola di Pubblicistica alla Città universitaria—Facoltà di Giurisprudenza di Roma—nell 'anno accademico 1956–1957.
92. Deleted on revision.
93. BALBIS, BRUNO. Le développement de la documentation en Italie. *Revue de la documentation,* v.21.2, June 1954. pp. 53–59.
94. MORGHEN, RAFFAELLO. Preside, Scuola speciale per Archivisti e Bibliotecari, letter dated December 11, 1957, from Rome.
95. BOURGEOIS, letter, *op. cit.*
96. WILSON, in Ashworth, ed., *Handbook*…. *op. cit.,* p. 354.
97. SOKOLOV, V., GARR, M., KULETZINA, A. Improving the leadership of technical libraries (in Russian). *Bibliotekaŕ,* no.2, 1953. pp. 20–21.
98. COLLISON. Fourth…. *op. cit.,* p. 108.
99. WHITBY, THOMAS. Head, Cyrillic Group, Union Catalog, Library of Congress, communication during discussion October 8, 1957.
100. *Ibid.*
101. MIKHAILOV, A.I. letter dated December 12, 1957, from Moscow.
102. New institute of librarianship and bibliography (Moscow). *Unesco Bulletin for Libraries,* v.11.7, July 1957. p. 179.
103. Cursos para la formacion tecnica de archiveros, bibliotecarios y arqueologos, 1957–1958. Direccion general de archives y bibliotecas, Madrid.
104. ESTEVE, FRANCISCO. letter dated January 4, 1958, from Madrid.
105. MILAĈIC, DUŜAN. letter dated November 16, 1957, from Belgrade.
106. *Ibid.*
107. JOEL, I. letter dated December 27, 1957, from Jerusalem.
108. DELOUGAZ, NATHALIE. Library training in Israel. *Unesco Bulletin for Libraries,* v.11.4, April 1957. pp. 82–83.
109. COLLISON. Fourth…. *op. cit.,* p. 100.
110. CARNOVSKY, LEON. *Report on a Programme of Library Education in Israel.* Unesco, Paris, 1956.
111. WORMANN, CURT. The scientific libraries of Israel. *Unesco Bulletin for Libraries,* v.8.1, Jan. 1954. pp. 1–5.
112. COLLISON. Fourth…. *op. cit.,* p. 18.
113. VLEESCHAUWER, H.J.DE. The University of South Africa and its department of library science. *Mousaion,* no.12, 1956.
114. Pretoria. University of South Africa. Department of librarianship. *Unesco Bulletin for Libraries,* v.9.10, Oct. 1955. p. 220. review of the new *Mousaion*.
115. MEWS, HAZEL. Chief Information Officer, Library and Information Division, South African Council for Scientific and Industrial Research, letter dated November 1, 1957, from Pretoria.
116. ——*Books are tools.* C.S.I.R. Library and information division, Pretoria, 1951.

117. C.S.I.R. Library and information division. *Summary of activities,* n.d. mimeo. WILSON, in Ashworth, ed., *Handbook…. op. cit.,* pp. 348–349.
118. MALCLES. *op. cit.,* p. 174. JOHNSON, A.F. “Library training programs.” *Australian Library Journal,* v.2.1, Jan. 1953. pp. 3–5.
119. WILSON, in Ashworth, ed., *Handbook…. op. cit.,* pp. 346–347.
120. JOHNSON, *op. cit.*
121. WHITE, H.L. The Commonwealth National Library of Australia. *Unesco Bulletin for Libraries,* v.7.10, Oct. 1953. pp. E124–127.
122. BRIET. *Enquiry…. op. cit.,* p. 32.
123. PETERSEN, E.N., Acting Head, Libraries Division, Unesco, typed bibliography of articles in *Unesco Bulletin for Libraries* accompanying letter dated October 18, 1957, from Paris.
124. COLLISON. Fourth…. *op. cit.,* p. 33. MALCLES. *op. cit.,* p. 38,
125. UNESCO. LIBRARIES DIVISION. *Bibliographic Newsletter,* v.6.2, 1957. pp. 40–41. Brazilian courses in bibliography. *Unesco Bulletin for Libraries,* v.11.8, Aug. 1957. pp. 218.
126. COLLISON, Third…. *op. cit.,* pp. 31.
127. *Ibid.,* p. 32.
128. LITTON, GASTON. Inter-American school of library science, Colombia. *Unesco Bulletin for Libraries,* v.11.8, Aug. 1957. p. 198.
129. Mexico: scientific and technical documentation centre. *Bibliographic Newsletter* of the libraries division of Unesco, v.1.2, 1952. pp. 18–19.
130. MALCLES. *op. cit.,* p. 162.
131. COLLISON. Third.... *op. cit.,* p. 35.
132. *Ibid.*
133. SAHA, J. letter dated November 27, 1957, from Calcutta.
134. COLLISON. Fourth…. *op. cit.,* p. 51. KEIO-GIJUKU UNIVERSITY. FACULTY OF LITERATURE. JAPAN LIBRARY SCHOOL. *Announcement catalogue*. The University, Tokyo, 1958. (also personal experience)
135. A.L.A. B.E.L. Standards for accreditation, presented by the ALA/BEL, and adopted by the ALA council Chicago July 13, 1951. *ALA Bulletin,* v.46.2, Feb. 1952. pp. 48–49. see also ALA/BEL S*tatement of interpretation to accompany standards for accreditation*…. ALA, Chicago, 1952. (Board of Education for Librarianship)
136. U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE. OFFICE OF EDUCATION. *List of 563 institutions of higher education in the United States announcing courses in library science and/or bibliography.* Prepared by Willard O.Mishoff, Specialist for College and Research Libraries, Library Services Branch, July 30, 1957.
137. Library schools, in *Canadian Almanac and Directory for 1954,* edited by Beatrice Logan. Toronto, Copp Clark, 1954. p. 450.
138. MISHOFF, WILLARD O. Undergraduate programs of library education: a current summary. *Higher Education,* v.14.1, Sept. 1957. pp. 3–7.
139. Library schools, in *Canadian…. loc. cit.*

140. Home-study correspondence courses, in *Lovejoys Vocational School Guide*. N.Y., Simon & Schuster, 1955. Chapt. 14, pp. 193–201.

141. Subcommittee on course content of the Subcommittee on special library education of the Joint committee on education for librarianship of the Council of National Library Associations; Dr. Baer is Chief Librarian of the National Housing Center, Washington, D.C. Questionnaire letters were sent to the accredited schools listed in Appendix 2, and a preliminary report was made early in May 1957; the final report of the subcommittee probably will have appeared by the time the present paper is published, and may, of course, have reached different conclusions. The results given herein are based on study of the announcement catalogs of the schools surveyed, and on perusal of the individual returns from the survey for which permission was obtained from Dr. Baer and from each school reporting, to all of whom the writer is very grateful.

142. see *ALA Bulletin,* v.51.11, Dec. 1957. p. 888, footnote, (also Appendix 2)

143. LLOYD, GWENDOLYN. Survey of study facilities for foreign library school students in the U.S. *Special Libraries,* v.45.1, Jan. 1954. pp. 7–12. Supplement appeared in *Special libraries,* v.45.9, Nov. 1954. pp. 384–385.

144. MORLEY, LINDA. Special library education in the United States and Canada. *Journal of Documentation,* v.3.1, June 1947. pp. 24–42.

145. SHERA, JESSE. Training for specials; the status of the library schools. *Special Libraries,* v.28, Nov. 1937. pp. 317–321.

145a. Check H.W.Wilson Co.'s *Library Literature* for professional publications, etc.

146. Letters were sent to the presidents of the 31 SLA chapters in October 1957 asking what each chapter is doing in the area of education or training for special librarianship; no follow-up letters were sent. Seventeen reports were received by the end of January 1958; two more were promised but did not come. Results given herein are based on these returns; see Appendix, 3.

147. HARPER, SHIRLEY, and KIENTZLE, ELIZABETH. Special library problems: Illinois chapter education program. *Special Libraries,* v.44.6, July–Aug. 1953. pp. 236–242. Excellent description of course, its problems, and its evaluation. “During the last two years, it was found that, despite considerable expressed interest, actual participation was very small. Very likely another ‘cycle' of interest in formalized education programs may come in a few years.”—Illinois. Comments from other chapters were similar but less direct.

148. RECRUITMENT & TRAINING COMMITTEE. NEW YORK CHAPTER. SPECIAL LIBRARIES ASSOCIATION. *Directory of Training Opportunities for Special Librarians in Metropolitan New York.* New York, 1957. 19 pp. Attention is also directed to the local Adult Education Council, Vocational Advisory Service, and public library's Readers' Advisers Office.

149. e.g., Professional training for special librarianship: panel discussion. New York *Chapter News,* v.28.3, Feb. 1956. p. 5 (announcement and reason for having). KINGERY, ROBERT. S.L.A. N.Y. chapter Recruitment & training committee report. New York *Chapter News,* v.29.4, April 1957. pp. 10–11. Well-thought-out statement of principles, needs, and limitations of special library education.

Special training *vs.* library training for special libraries. Texas *Chapter Bulletin,* v.7, Nov. 1955.

150. New library course due. *New York Times,* May 17, 1953. (at Queens College, suggested by the New York Chapter) Ballard school course announcements periodically in New York *Chapter News;* elementary reference work, subject filing and indexing, elementary cataloging, are among those offered the year around. The Ballard School is in the YWCA and has been sponsored and developed by the New York chapter; certificates are given to those who complete the courses. Two members of the Oak Ridge chapter "have been trying to establish a cooperative program with our company for the training of special librarians. The company has such a program for training engineers and seems sympathetic to the idea of including librarians in it." "A long range hope of our chapter is to cooperate in some library course or workshop at the University of New Mexico. Because of the lack of a library school in the area we feel the need for something like that…. In our libraries at the Los Alamos Scientific Laboratory we are experimenting with a reference course for technical library work which is aimed at refreshing the minds of our science librarians…. In a few years, after our chapter members have gained experience and confidence with such local ventures, we hope to take an active part in library training in the state."—Rio Grande, one year old. "Washington, D.C., chapter-sponsored U.S. Department of Agriculture Graduate School courses (current): Introduction to cataloging and classification, Principles of library organization, Basic reference service and reference tools, Introduction to bibliographic science, Law librarianship, Introduction to map library techniques, The principles of physical science (survey course in fundamental concepts)." "We feel that it is so vital that we have made 'Education for special librarians' our 1957–58 Chapter project. Our approach to this problem has been to study the present library training facilities in the state…(few courses offered, few professors have knowledge of needs)…. Therefore, we felt one of the functions of our chapter could be to formulate our educational needs and to attempt to develop a mutually cooperative program with the schools to meet these needs…. We have appointed to the project committee the subject specialist, that is, a chemist with no library training, and conventional library school graduates without the subject specialization. In addition we have asked the key library schools to send representatives to our meetings to act as advisors. We have also asked the president of the Library Education Counsel of the state Library Association to act as an advisor. In this manner we can openly discuss our educational needs with these people and, therefore, we can generate interest in our problems at the library school level…. We feel that this is a long-term program. We do not anticipate a solution this year or even next year. We have, however, made some progress…(extension courses, Saturday classes) …. We are actively participating in the state Library Association Library Education Counsel and the Advisory Counsel of the Graduate School of Library Science at the University (to resolve the problem of poor communication

between the special and the ordinary librarians)….”—Texas, very big and very active, especially in the area of special libraries.

151. HENKLE, HERMAN. *Dissemination of information for scientific research and development with special reference to the work of Research information service, The John Crerar Library*. The Library, Chicago, 1954.

152. CRANE, E.J. The training of chemists for abstracting and indexing, in *Training of Literature Chemists,* Advances in Chemistry series no. 17. American Chemical Society, Washington, D.C., 1956. pp. 16–21.

153. Check list of proceedings for new reference personnel. Battelle Memorial Institute, Columbus, processed, one page.

154. DUGGAN, MARYANN. Technical Librarian, Magnolia Petroleum Company, letter dated November 19, 1957, from Dallas.

155. HOFFMAN, THELMA. Chief Librarian, Shell Development Company, letter dated December 17, 1957, from Emeryville, California, also enclosed, Shell's statement giving descriptions and personnel requirements of non-laboratory research positions in technical files, patent library, and technical library.

156. DAHL, RICHARD. Professional development program. *Library Journal,* v.79.21, Dec. 1, 1954. pp. 2280–2283. Benefits and methods are given, for all libraries. SHANK, RUSSELL; BEHYMER, E.HUGH; METCALF, KEYES. Staff participation in library management. *College and Research Libraries,* v.18.6, Nov. 1957. pp. 467–478.

157. WESNER, JEAN. Training of non-professional staff. *Special Libraries,* v.46.10, Dec. 1955, pp. 434–440. Those with no formal library science education and no experience in the field. Needs and methods demonstrated. MCNEAL, ARCHIE. “Ratio of professional to clerical staff.” *College and research libraries,* v.17.3, May 1956. pp. 219–223. Comments at some length on need for training of non-professionals. MULCAHEY, J.H. Training special library assistants. *Special Libraries,* v.48.3, March 1957. pp. 105–108.

158. CHAFFEE, RANDOLPH. The engineering library. *Machine Design,* v.24.9, Sept. 1952. pp. 110–127. “To attract more enrollments in special library courses, industry might offer scholarships or part-time employment in technical libraries to post-graduate students.” Literature chemists scarce, too. *Chemical and Engineering News,* v.33, April 11, 1955. pp. 1518–1519. Industry “has been forced to provide on-the-job training.”

159. NORWOOD, M.L. Survey of current library in-service training practices. Thesis, University of North Carolina, 1957. WOODRUFF, ELAINE. In-service training for government librarians. *Special Libraries,* v.44.2, Feb. 1953. pp. 48–52. Program needs and schedule given, adaptable to all kinds of libraries.

160. GRUNWALD, WILHELM. “Der Spezialbibliothekar: Aufgaben, Auswahl, Ausbildungsvorschläge.” *Bibliothek, Bibliothekar, Bibliothekswissenschaft; Festschrift Joris Vorstius*…1954. pp. 182–191. Interesting proposed training program. FILL, KARL. Thesen zur ausbildung der dokumentalisten. *Nachrichten für Dokumentation,* v.5.1, March 1954. pp. 28–32. Stresses needs and then courses.

PRESSL, LISSI, and SCHMOLL, GEORG. Einige Gedanken und Materialien zur Ausbildung von Dokumentalisten in der DDR. *Dokumentation,* v.4.2, March 1957. pp. 31–37. Includes a 20-category lesson plan and a bibliography.

161. MALLABER. *op. cit.* MACKIEWICZ, ELIZABETH. Education for special librarianship in Great Britain. *Aslib proceedings,* v.5.4, Nov. 1953. pp. 286–292. Training for librarians and documentalists. *Aslib proceedings,* v.6.2, May 1954. pp. 117–119. Points out differences between librarians and information officers. FARRADANE, J. Information service in industry. *Research,* v.6, Aug. 1953. pp. 327–330. Also discusses training of the information officer. KAY, A.G. Qualifications for special library work. *The Engineer,* v.198, July 23, 1954. pp. 130–132. HARRISON, J.C. Special librarianship and the library schools. In Library Association, Reference and special libraries section, *Proceedings of the annual conference 1956.* London, 1957. pp. 23–28. Training of special librarians and information officers. *Aslib proceedings,* v.9.4, April 1957. pp. 99–100. BENGE, R.C. The place of literature studies in library education. *Journal of Documentation,* v.13.3, Sept. 1957. pp. 147–151. Stresses importance of subject literature study to both librarians and information officers, in schools. DAIN, N.E. review of and comments on Harrison's paper (above). *Library Association Record,* v.59.10, Oct. 1957. pp. 347. HARRISON. comments and reply to Dain (above). *Library Association Record,* v.59.11, Nov. 1957. p. 375. HOLMSTROM, J.E. Observations on the training of information officers. Unesco Department of natural sciences, 30 September 1953. Reference 250/3930. mimeo. Gives definitions, functions, and suggested training syllabus. Similar to point of view of Farradane (above). Deplores lack of effect of previous appearance of observations (27 September 1948) on Library Association syllabus drafting.

162. ASHEIM, LESTER, ed. *The core of education for librarianship.* ALA, Chicago, 1954. Suggests seven "areas of the core," and what might be included in library training at the undergraduate level and in special subject fields. ——Education for librarianship. *Library Quarterly,* v.25.1, Jan. 1955. pp. 76–90. Review of 1931–1955 and of beginning of demand for special library training. BRODMAN, ESTELLE. Whither education for medical librarians? *Stechert-Hafner Book news,* v.8.6, Feb. 1954, pp. 61–62. DISBROW, MARY. Impressions of the course in medical libraries at Emory University. *Bulletin Medical Library Association,* v.41.3, July 1953. pp. 277–282. Included lectures on departments of medicine by medical men. Education for special librarianship. *Library Quarterly,* v.24.1, Jan. 1954. pp. 1–20. Includes suggested education in science-technology and medicine library work. LANCOUR, HAROLD. The training of the special librarian in the U.S. *Aslib proceedings,* v.5.4, Nov. 1953. pp. 271–275. Includes development of pattern. LEONARD, RUTH. Education for special librarianship. *Special Libraries,* v.41.5,

May–June 1950. pp. 157–159. Description and discussion of Simmons College program. POSTELL, WILLIAM. Education for medical librarianship. *Special Libraries,* v. 48.5, May–June 1957. pp. 186–188. Development and present needs. SHERA, JESSE. Education for librarianship—an integrated approach. *ALA Bulletin,* v.48.3, March 1954. pp. 129–130, 169–173. ——*et al.,* ed. *Documentation in Action.* Reinhold, New York, 1956. *passim.* ——Librarians' new frontier. *Library Journal,* v.82, Jan. 1, 1957. pp. 26–28. ——Research and developments in documentation. *Library Trends,* v.6.2, Oct. 1957. ——The training of the chemical librarian: a challenge and an opportunity. *Special Libraries,* v.47.1, Jan. 1956. pp. 8–16. TAUBE, MORTIMER. Implications for professional organization and training, in *The Communication of Specialized Information* edited by Margaret Egan. University of Chicago Graduate Library School, 1954. WATERS, EDWARD. Special library education. *Library Trends,* v.1.2, Oct. 1952. pp. 244–255. Whole issue devoted to special libraries. ——Training for special librarianship. *Special Libraries,* v.47.9, Nov. 1956. pp. 393–399. Gives definitions, deficiencies, training programs, and progress.

163. BRIET. *op. cit.* (ref. 22) VAN DIJK. *op. cit.* (ref. 36) PRZELASKOWSKI. *op. cit.* (ref. 87) VLEESCHAUWER. *op. cit.* (ref 113)

164. A.L.A. B.E.L. Standards…. *op. cit.* (ref. 135)

165. Library education division. *ALA Bulletin,* v.51.11, Dec. 1957. p. 858. See also reports on activities of Joint Committee on Library Education, e.g., in Association of American Library Schools *Report of meeting,* Chicago, January 28, 1957, pp. 59–62, in which Dr. Carl M.White explores the whys and wherefores of the committee trying to pin down its responsibilities, if any. *AALS Newsletter,* January and July, frequently contains news and comments on instructional programs in the area of special librarianship. LED *News Letter,* including its *Teachers section issue,* frequently contains news and reports of activities in all areas of education for librarianship, surveys, new courses, committee reports, and so forth.

166. MERTON, ROBERT. The functions of the professional association. *American Journal of Nursing,* v.58.1, Jan. 1958. pp. 50–54.

167. *Ibid.,* p. 52.

168. COBLANS, HERBERT. New methods and techniques for the communication of knowledge. *Unesco Bulletin for Libraries,* v.11.7, July 1957. pp. 154–175. KYLE, BARBARA. Current documentation topics and their relevance to social science literature. *Revue de la documentation,* v.24.3, Aug. 1957. pp. 107–117. The present state of "Research in librarianship" is examined and discussed from all points of view in the October 1957 *Library Trends;* the entire issue is devoted to the subject of research in librarianship.

APPENDIX 1 ASLIB TRAINING COURSES

The Education Committee has had under consideration means of extending Aslib's programme of short training courses and has agreed that courses on the following subjects are desirable:

Organization of Science and Technology

Good relations with users and presentation of information to management and administration

Acquisition, handling and exploitation of periodicals

Advanced classification and indexing

Production of library and information department publications

Mechanical information retrieval

Work with technical reports

Patents and patent law.

Work is proceeding on the organization of these courses.

APPENDIX 2 SURVEY OF ACCREDITED LIBRARY SCHOOLS

	Special libraries			*Literature courses*				*Courses offered*			
Short name of school	*Prog.*	*Cour.*	*Semr.*	*S-T*	*Doc*	*Med*	*Oth*	*U-g*	*Ext*	*Eve*	*Sum*
Atlanta				e				x			x
California				e	x						x
Carnegie Tech	s-t	s-t		e	s						
Catholic University		x						x			x
Chicago		x		r				x		x	x
Columbia	s,m			r,e	x	x	i,p	x			x
Denver		x		r							x
Drexel Institute	x	x		e							
Emory	c					x		x			x
Florida State		x	x	e				x	x		x
Illinois				e		x	a,b				x
Indiana				r				x			x
Kentucky								x			x
Louisiana State		x						x			x
McGill		x			x			x			
Michigan		x		e	s	x	i	x	x		x
Minnesota		x		r				x		x	x
North Carolina			x	e		x		x			x
Oklahoma	x			r				x			x
Peabody	m			r				x			x
Pratt Institute		x		e							x
Rosary				r						x	x
Rutgers			x	e	x			x	x	x	x
St. Catherine				e				x		x	x
Simmons	x	x	x	e				x		x	x
So. California		s-t		r,e				x	x		x
Syracuse				r				x		x	x
Texas State College for Women		x		r				x			x
Texas	x	x						x	x		x
Toronto		x						x			
Washington, St. Louis		x								x	x
Washington, Seattle				r				x			x
Western Michigan			x	e				x			x
Western Reserve		x		e	x		*			x	x
Wisconsin		x						x			x
Total 35	8	19 (15 add. 3)	5	13r 16e	5x 2s	5		26	5	9	31

a, agriculture; b, biology; c, chemistry; e, elective; i, indexing/abstracting; m, medicine; p, pharmacy; r, required; s, science; t, technology; U-g, undergrad; *, machine lit., etc., and Center for Documentation and Communication Research. Cour., course; Doc, documentation; Eve, evening; Ext, extension; Med, medical; Oth, other; Prog., program; S-T, science-technology; Semr., seminar; Sum, summer.

APPENDIX 3 SURVEY OF SLA CHAPTERS

	Education/training programs outside chapter liaison with nearby library school						*Programs inside chapter (held during 1956–1957)*			
Chapter	*talks*	*teach*	*tours*	*cnslt*	*w/stud*	*other*	*none*	*wkshp*	*meth*	*lect*
Georgia	x		x	x	x			x		
Illinois	x	x			x	conf.			x	
Louisiana		x			x		x			
Michigan		x			x	conf.		x		
Montreal	x						x			
New Jersey	x	x	x		x			x		
New York	x	x	x		x	1			x	
Oak Ridge						2			x	x
Oklahoma							new			
Pittsburgh		x		x				x		
Puget Sound	x		x		x				x	x
Rio Grande	x			x			new			
San Francisco Bay	x		x	x	x				x	
Texas				x		3		x		
Toronto		x		x	x			x		x
Washington, D.C.		x				4	x			
Wisconsin										x
Totals 17	8	8	5	6	9		5	6	5	4

1. Instrumental in establishing Ballard School library courses, and courses at Queens College, New York.
2. Establishing cooperative training program within a company.
3. Active program with industry, library associations, schools.
4. Instrumental in getting courses set up in USDA graduate school.

cnslt, consult; conf, conference; lect, lecture; meth, methods; USDA, United States Department of Agriculture; w/stud, with students; wkshp, workshop.

Training the Scientific Information Officer

A.B.AGARD EVANS and J.FARRADANE

ABSTRACT. The function of a scientific information officer is basically different from that of a librarian. Documentation has been defined as the recording, organization, and dissemination of knowledge; the information officer is deeply concerned with all three. The information officer has a critical function; his task is synthesis of information, its predigestion and the addition of background information. The training therefore requires not only the inclusion of a variety of subjects not included in a librarian's training, but a fundamentally different approach and emphasis on what might appear to be common ground.

A syllabus is presented for a post-graduate course of training for a student who is already a subject specialist. Stress is laid on the problem of human communications in the efficient application of research; presentation of information at all levels of industry; abstracting, translating, editing; compilation of surveys; reproduction techniques. It includes indexing, classification, and work organization, but approached from a different angle than that of the librarian.

The course envisaged covers 166 hours of lectures plus practice. But the equipment of a competent information officer requires experience of research, laboratory, or practical work. It is therefore to be preferred that the course should not be consecutive to or concurrent with academic studies, but rather be available as a part-time or long-vacation course for experienced students.

THE FUNCTION OF A SCIENTIFIC INFORMATION OFFICER

The function of a Scientific Information Officer is basically different from that of a librarian. Documentation has been defined as the recording, organization, and dissemination of knowledge. The librarian is mainly concerned with organization; the information officer is deeply concerned with all three.

In a research establishment, one of the functions of the information officer is

A.B.AGARD EVANS Chief Librarian, Ministry of Works, London.

J.FARRADANE Scientific Information Officer, Tate and Lyle Research Laboratories, Keston, Kent, England.

the responsibility for assembling, organizing, and circulating the information required by the establishment. This function is that of the librarian with an intimate knowledge of the subject and the particular fields of interest of the research staff. But this work is only a part of his duties. He is responsible for ensuring that the work of the research staff is properly recorded and published, so that its use is maximised; that is to say, ensuring that it is systematically and adequately reported to facilitate accessibility of the information both currently and in future; and also that it may be fed into the world pool of information, where possible, so that it may reach other centres with the greatest certainty. He is in effect a publisher with a specialist responsibility.

Moreover, he has the responsibility for critical selection and predigestion of material required by the research staff both currently and in response to specific inquiry. This may take the forms of preparation of abstracts of papers, digests of information from several sources, translations, compilation of data sheets or information sheets; all of which is based upon recorded information. It is not his place to usurp the critical judgment of the research worker, but to prepare the material for the research worker's judgment in the same way as a solicitor compiles a brief for the barrister-at-law.

His answers to an inquiry will frequently go beyond that of the Reference Librarian: his task is not fulfilled by presenting a number of references or the actual material with the pages marked: he must prepare an essentially complete, accurate and up-to-date digest of available information, together with selected references for fuller study. The answer must be adjusted according to the context in which the information is required and the intellectual level or occupation of the inquirer.

Finally, his intimate knowledge of the work of the establishment and his wide range of reading and study put him in a unique position to help in the cross-fertilization of ideas, to become aware of gaps in knowledge and to make suggestions for future programmes.

It was the view of the Royal Society's Information Conference of 1948 that the Information Officer should have equality of training and status with his colleagues in research. The information section should be a major section in a research establishment and the Information Officer should rank equal with the Directors of the Research Divisions.

What has been said above in connexion with a research establishment applies to a greater or less degree to an industrial organization, a Government Department or business where the scientific or technological aspects predominate. By analogy, also, it could apply in other specialist fields with sociology or economics replacing the term science.

We offer no apology for the above brief exposition of a scientific information officer's function, which, though well-known is apt to be overlooked during controversy.

QUALIFICATIONS

It follows logically that apart from the essential *geist* of aptitude, temperament and vocation for this ‘peculiar trade', the education and training for the work must comprise: *(a)* scientific qualifications of a good standard; *(b)* reading knowledge of three major languages; *(c)* training in the technique of technical information (TTI).

At first sight there would appear to be a large measure of common ground between the training of a librarian and TTI. On this assumption a joint committee of the Library Association and Aslib (Association of Special Libraries and Information Bureaux), strove earnestly between 1954 and 1957 to evolve a syllabus which would, in whole or in part, serve both purposes. No real progress was made, however, towards a syllabus or courses suitable for persons engaged in those aspects of information work quite distinct from librarianship.

The fact emerged more and more clearly that not only does TTI include a variety of subjects not relevant in a librarian's training, but that it also requires a fundamentally different approach and emphasis on what might appear to be common ground.

SYLLABUS

A syllabus is presented for a post-graduate course of training for a student who is already a subject specialist. The syllabus is not a product of spontaneous genius, but owes much to the deliberations of the Aslib Education Committee, which, however, is not officially responsible for it as presented here. The course covers 166 hours of lectures plus practice. Since the equipment of a competent information officer requires experience of research, laboratory, or practical work, it is to be preferred that the course should not be full-time or concurrent with academic studies, but rather be available for part-time or evening studies, possibly culminating in a vacation seminar of a month.

It will be noted that stress is laid on the problem of human communications in the efficient application of research; presentation of information at all levels of industry; abstracting, translating, editing; compilation of surveys; reproduction techniques. It includes indexing, classification, and work organization, but approached from a different angle from that of the librarian.

SYLLABUS

1. Human communications: some problems in the efficient use of the results of research in science and scholarship
2. Inquiry techniques
3. Sources of information

(a) general reference sources
(b) materials and sources in
   either (i) science and technology
   or (ii) social sciences

4. Cataloguing, indexing, and classification
5. Presentation of information

(a) abstracting and précis writing
(b) principles of translating
(c) editorial work and report writing
(d) compilation of selective bibliographies, surveys, and special reports
(e) dissemination of information, including

(i) internal and external publications
(ii) personal liaison
(iii) answers to inquirers

6. Administration

(a) personnel, work organization, budgeting, and costing, planning, equipment, and layout
(b) organization of internal records
(c) book selection, loans, exchanges

7. Document reproduction
8. Copyright law.

Notes on Syllabus

1. Course 1. (8 hours) consisting of

1. Introduction
2. Growth of language difficulties, including

(a) relation of concept and term
(b) source and nature of transliteration problems
(c) standardization problems

3. Proliferation of written records

(a) scientific
(b) sociological

4. *(a)* national listing
   (b) international subject listing
5. Inadequate publication
6. Accessibility of material

(a) library development
(b) non-library sources

7. The economic factor
(a) relating to publication
(b) affecting the nature and extent of service to research

2. Course 2. (3 hours) consisting of

1. Assessment of inquiries
2. Factual information
3. Indefinite and research inquiries

3. Course 3(a). (12 hours+practical) consisting of

1. Book trade	1 hour
2. Government publications	2 hours
3. U.N. and foreign government publications	2 hours
4. Maps	1 hour
5. Periodicals indexes	1 hour
6. Abstracts	3 hours
7. Bibliographies of bibliographies	1 hour
8. Guides to current research	1 hour

4. Course 3(b)(i). (48 hours+practical work) to include the following:

   Periodicals, house journals, etc., and indexes to them; preprints and offprints; catalogues of periodicals; abstracts; book reviews; reviews of progress; current and standard bibliographies; monographs series; collections of tables and data; Handbücher; technical dictionaries—single language and others; standards; patents and their use; book selection

   Libraries; interlibrary co-operation, national and international; societies; institutions; trade associations; research associations; research stations; chambers of commerce; government departments; intergovernmental organizations; other international organizations

5. Course 3(b)(ii). (48 hours+practical work) to include the following:

   Newspapers—indexes, cuttings, press summaries; monitoring reports of broadcasts; periodicals—indexes, abstracts, offprints, etc; surveys—reports of sample surveys, national, local, government and commercial Monographs; dictionaries, and glossaries.

   Statistical sources—censuses and other basic national statistics, statistical abstracts, statistical publications of government departments, international statistical publications; legal sources—national and international Bibliographical sources—publications of the International Committee for Social Science Documentation, etc., records of current research including theses, trend reports, reviews, library catalogues and acquisition lists, commercial bibliographies

   Centres of documentation—handbooks and guides to centres; inter-library co-operation—joint catalogues of periodicals, etc., societies and institutions

   Special problems of social science documentation:

1. Area studies
2. Problems connected with a new orientation of disciplines

(a) terminology and, therefore, classification
(b) importance of sources in other disciplines making past or present contribution to social science studies

6. Course 4. (48 hours) to be based on following division:

1. Author treatment

2. Subject treatment

(a) in catalogues

(i) alphabetical subject headings

(ii) general classifications

(b) in indexes (distributed and collective entries, with reference to special classification, etc.)

7. Course 5 as follows:

(a) 24 hours
(b) 1 hour
(c) 3 hours to include proofing, preparation of MSS,
(d) 3 hours
(e) 3 hours

8. Course 6. (9 hours) as follows:

(a)	1. personnel, work organization	1 hour
	2. budgeting and costing	2 hours
	3. planning, equipment, and layout	2 hours
(b)	organization of internal records	2 hours
(c)	book selection, loans, exchanges	2 hours

9. Course 7. (3 hours) (including visits)
10. Course 8. (1 hour).

Training for Scientific Information Work in Great Britain

B.I.PALMER and D.J.FOSKETT

ABSTRACT. The demand for the services of information workers has been met in Britain in various ways, sometimes by secondment from laboratory work and sometimes by extension of the librarian's function. As many of the Library Association's members are engaged in special libraries, the Library Association has a natural interest in their training and qualification. The basic techniques of the librarian and the information officer are the same, and the Association's examination system recognises this. The development of the Library Association's syllabus since World War II has been directly conditioned by the development of information work, and examples of this are given. The growth of a teaching system correlated with the Association's syllabus and the extension of this system to specialised applications of librarianship are noted. Aslib's continuing interest in the education of information officers is also shown, although they have decisively rejected a proposal to establish a separate register of information officers, in the interests of professional solidarity. It is shown that experience leads to the conclusion that the British system of qualification is a valuable and flexible one, and is worthy of study by others, as it brings together workers in all fields of information service, and each gains in strength from this association.

There is no prescribed course of training available in Great Britain for the scientific information officer as such. The demand for such workers has been met in various ways as the need arose within various organisations. In some cases, scientists have found themselves assuming responsibility for this work on behalf of their colleagues, and their subsequent appointment as Information Officers has been a *de jure* recognition of a *de facto* circumstance. In other cases, where the organisation concerned already had a library service functioning, the work of the information officer has been regarded as the natural dynamic extension of that of the librarian, and he has carried out the work, with or with

B.I.PALMER The Library Association, London.
D.J.FOSKETT The Institute of Education, London University, London.

out the addition of the title of Information Officer to his existing title. Yet other cases have occurred where a delimitation of functions has enabled both a librarian and an information officer to work together, the first as a provider and disseminator of information, and the second as a trained user of information services acting as an agent for his scientific colleagues (1). Some would go further and declare that there is no difference at all between a special librarian and an information officer. In discussing a paper at the 20th Aslib Conference held in London, September 1945, Mr. A.B.Agard Evans said:

Mr. Simons was in an elementary but fundamental error in trying to distinguish between a special librarian and an information or intelligence officer. They were precisely the same animal, whatever the local term…. Roughly 99% of information was recorded either in print, typescript, photographs or files. A special librarian collected, classified and indexed the literature and answered questions by means of it; an intelligence officer, on the other hand, found the information he required by consulting the literature he had previously collected and filed. Both used outside sources freely, knowing the limitations of their own (2).

For sixty years the Library Association has been the qualifying body for librarians in the United Kingdom, and for nearly thirty years it has provided, as we shall show, some form of examination in the documentation of science and technology. In recent years a great deal of study and discussion have taken place, and the syllabus of examinations has been continuously modified and improved in this respect.

Many members of the Library Association who are qualified librarians ("Chartered Librarians") are directing or employed in special libraries devoted to the organisation and dissemination of scientific and technological information. Careful examination of the kinds of work they engage in, and comparison with that of librarians working in other fields, leads us to conclude that there are certain basic techniques common to workers in all kinds of libraries, and that these are susceptible to development to meet the varying needs of different kinds of libraries. These techniques are (1) organisation of knowledge through schemes of systematic classification, through catalogues and through indexing, (2) the dispensing of information in response to enquiries, through enquiry techniques and bibliographical research, (3) the dissemination of information in anticipation of needs through a study of the work of the research staff, the preparation of displays, bibliographies, abstracts, and so on, up to and including full-scale literature surveys, (4) the organisation of an integrated library and information service, to deploy staff and resources to achieve these ends.

On the basis of a satisfactory examination result after study of these techniques and practical in-service training, a general qualification in librarianship is granted, that is, election to the Register of Chartered Librarians in the category of Associate, or Chartered Librarian.

Not unnaturally, the examinations have tended always to pre-occupation with that branch of the library service which was developing at the time of the formulation of the syllabus. The great extension of scientific information work since World War II, and the corresponding increase in the numbers of workers in research and industrial libraries, have brought about a concentration of attention on those parts of the syllabus which have ceased to cater adequately for these workers.

Very soon after World War II, despite the very recent introduction of a new syllabus of examinations in 1946, the Library Association set up a syllabus subcommittee to review the syllabus in the light of developing post-war conditions, and invited representatives of all interests in the library and information field, including Aslib. The result was an improved syllabus (3) effective from 1950 onwards, making greater provision for specialised alternatives in the Final Examination, which is subsequent to the general (Registration) examination which qualifies Associates. Some idea of the development in the last 25 years can be gained from Tables 1 and 2.

It will be noted that one of the subjects listed under Table 1, column 3, has 1959 as the date of implementation against it. This subject "The Presentation and Dissemination of Information" was first projected about 1952–1953, when it was the subject of discussion by the Reference and Special Section of the Library Association. It was adopted as part of the syllabus in 1957, and comes into effect after a statutory 2-year period which allows teaching bodies time to provide the appropriate courses. There follows an outline of the syllabus for "Presentation and Dissemination of Information."

Presentation of ideas, including composition, style and language, readership, choice of material. Types of publications: reviews, house journals, annual reports, etc. Methods of reproduction and printing of foregoing. Editing, including law of libel. Preparation for the press. Copyright in dissemination. Abstracting and the form of abstract journals, preparation of reports and publicity materials. Collation of abstracts with originals. The principles and practice of indexing in special libraries, and the recent developments in mechanical and electronic methods.

Some specimen questions based on this syllabus are as follows.

1. How would you ensure that confidential information was widely and rapidly disseminated through an organisation of 500 people in three establishments, a headquarters and two branches?
2. You are preparing an accessions list for a circulation of 1,000. What method of reproduction would you use, and why?
3. Draw up a set of instructions for the use of a body of abstract translators dealing with a chosen subject or group of subjects.

4. What is the current position in this country regarding *(a)* fair copying and *(b)* protection of patents?
5. What is meant by Prof. R.O.Kapp's term "functional writing?" Discuss its implications for the dissemination of information in a particular organisation.

TABLE 1 Examinations in alternative subjects available to librarians in specialised fields

Examination	*1930*	*1945*	*1950*
Registration	—	—	Literature of Science Literature of Social & Political ideas
Final	Literary History of Science Literary History of Economics & Commerce Indexing & Abstracting University & Special Library Administration	The Literature of Philosophy & Religion Social Science Science & Technology Fine Arts Music Medicine Advanced Classification & Cataloguing Administration of Special Libraries & Information Bureaux	The Literature & Librarianship of Philosophy & Religion Social Sciences Science & Technology (general) Mathematical & Physical Sciences Chemistry & Chemical technology Natural History & Biological Sciences, pure & applied Engineering & Building technology Fine Arts Music Medicine History & Archaeology Linguistics & History and Theory of Literature Administration of Special Libraries & Information Bureaux Advanced Classification and Cataloguing Presentation & Dissemination of Information (from 1959)

TABLE 2 Numbers of candidates taking the special alternatives in the Registration Examination

Examination	*1950*	*1957*
Literature of Science and Technology	10	44
Literature of Social and Political Science	24	57

In 1955, as a result of a resolution passed by the Aslib Conference at Nottingham in September 1953, proposals were adopted for the setting up of a joint committee of representatives of the Library Association and Aslib to consider ways in which the Library Association's Syllabus could be made to suit more fully the needs of research library and information workers. With certain points of difference between the two bodies unresolved, there was produced a suggested syllabus which sought to provide a core of general librarianship, somewhat less detailed than the present Registration Examination, and a series of alternative papers in special techniques and subject fields. The general formula for library studies in special subject fields reads as follows:

The standard works and editions. Sources of literature, e.g., societies and Government Departments. Special types of literature where applicable: “Handbücher,” monographs, reviews of progress, catalogues of specimens, log books, maps, tables, sound recordings, etc. Evaluation and selection of material. Policy of book selection as determined by clientele and cooperative provision. Methods of acquisition.

Bibliographical apparatus. Bibliographies of bibliography, guides to literature, standard bibliographies, library catalogues, bibliographies to books and articles, indexing and abstracting services, surveys, guides to libraries.

Treatment of the subject in the principal general classification schemes, and in any special schemes. Special problems of classification and cataloguing thrown up by the subject field.

Outstanding collections in the field, their contents, special features, and availability.

Exploitation of a library in the field. Assistance to readers, display, production of booklists, literature surveys. Dissemination of information.

Using this suggested syllabus (4) as a basis, the Library Association's own Syllabus Sub-Committee is now meeting regularly with the object of completely revising the existing syllabus, to bring it even more into line with current needs of libraries, including scientific and research libraries.

The existence of a syllabus of examinations leading to a qualification offering professional status has called into being a considerable teaching system, nationwide in its provision (3). The natural home for such teaching in Britain is the Technical College, which already has a tradition of tuition for external examinations, including university degrees. This institution is a flexible one, being willing to provide either full or part time courses, according to demand. It can use the services of practising librarians as part-time teachers, and can provide a course on any subject if more than a handful of students are forthcoming. There are now 9 such institutions offering full and part time courses in preparation for the Library Association's examinations, and a further 38 offering part time courses only. Additionally there are frequent part-time and specialist courses held by such institutions, using the lecturing staff assembled for teach ing in connection with the Library Association's examinations. Examples are those provided by the North Western Polytechnic, London, in the Literature of Science; and by Manchester College of Science and Technology in Special Librarianship and Information Work, and in Textile Information Work.

This last has been conducted in co-operation with Aslib, which has long concerned itself with the training of information officers (5): indeed, Aslib annually offers a short “first aid” course for junior information officers and workers in special libraries, and from time to time has also provided short refresher courses of about a week's duration for senior workers in the field. It intends to develop this side of its work considerably in the future; the last Aslib conference, however, decisively rejected a proposal to establish a Register of Information Officers, as a qualification gained by examination, or the equivalent.

This factual picture of the situation with regard to the education, training and qualification of those engaged on scientific information work in Britain is offered as an example of what can be done by professional associations to encourage the study of documentation techniques. All the experience of the Library Association (and of other British professional associations) goes to show that the offer of a definite qualification leading to professional status, based partly on an examination upon a published syllabus, has a stimulating effect upon education for the profession. If every country has not the same teaching facilities to hand, the system is flexible enough to lend itself to any circumstances. It should here be pointed out that the Library Association examines in many countries overseas the candidates preparing themselves by correspondence or private study. The correspondence courses of the Association of Assistant Librarians have done valuable and world-wide work in training students in out-of-the-way places for general library work. Professional associations in other countries might find such courses a useful addition to their own endeavours.

Our experience has convinced us that the profession of librarian and information officer is one, using the same kinds of materials and techniques, even if not always in the same way. We also believe that, contrary to what is sometimes maintained, there is no characteristic peculiar to scientific literature that necessitates a distinct profession in that field (6). The services given to scientists by librarians, information officers, literature chemists, etc., are no less needed in other fields of knowledge, and will without any doubt develop in a similar manner. Yet it is likely that the profession will never be great in numbers, and while we may recognise many variations in both the pattern of organisation and the particular subjects studied in libraries, we cannot feel that complete fragmentation in professional education is possible or desirable (7).

The pattern of qualification evolved over nearly three quarters of a century of discussion and practice by the Library Association has much to recommend it, since it brings together workers in all fields of information service, and each gains in strength from this association. We suggest that professional bodies concerned with documentation in other countries might well be able to benefit from this experience in establishing standards, and building up a system of education to meet them.

REFERENCES

1. The Library Association. Memorandum on the rank and status of special librarians and information officers, priv. circ., 1949.
2. The Association of Special Libraries and Information Bureaux. *Report of the Proceedings of the Twentieth Conference,* London, 1945, p. 60.
3. The Library Association. *Students' Handbook.* Annually, January.
4. The Library Association Record, March, 1957, pp. 97–99.
5. WILSON, L. The education of the information officer in the United Kingdom, in *Congrés international de bibliothèques et de centres de documentation,* vol. IIB, Nijhoff, La Haye, 1955.
6. FARRADANE, J. Professional education of the information scientist. *Ibid.*
7. STOKES, R.R. Education for librarianship. *Ibid.*

The ICSU Abstracting Board: The Story of a Venture in International Cooperation

G.-A.BOUTRY

1. HISTORY

On June 20, 1949, an International Conference met in Paris, at the seat of Unesco, Avenue Kléber. It had been organized by the Department of Exact and Natural Sciences of Unesco (Professor P.Auger, Director) and was convened to deal in problems connected with scientific abstracting. One hundred nine delegates attended; they represented either countries (24 nations had nominated representatives), international institutions or organizations. Out of the eleven Scientific Unions federated by ICSU, eight had sent delegates.

Most of the problems or difficulties of scientific abstracting were dealt with and were mentioned in the recommendations listed in the Final Act. Of these, three are of special interest to us today. One (recommendation 6) emphasized the importance of establishing good cooperation between existing scientific abstracting services and proposed the creation of coordinating committees:

…It is recommended that abstracting agencies cooperate for the improvement of their services by extending agreements for the exchange of abstracts and of original material for abstracting, and by agreeing on the subject field and services of each.

Another recommendation (No. 11) urged that all original articles appearing in scientific journals should begin or end with an author's abstract (synopsis), prepared according to a set of rules to be established later. The “Guide for the preparation of Synopses,” which had just been published by the Royal Society (U.K.), was cited as an excellent basis or starting point for the discussion of this matter:

…It is recommended that each issue of a scientific journal include synopses, in English or French at least, of all original articles contained in it; that the editor-in-chief of the journal accept responsibility for the adequacy of these synopses, whether or not prepared by the author; and that the journal state in each issue that the republication of all its synopses in whole or in part is authorized;….

G.-A.BOUTRY Secretary, International Council of Scientific Unions Abstracting Board.
The ICSU Abstracting Board is discussed further in a paper by B.M.Crowther, Area 2.

Recommendation 10 must be given here in full:

(10) *Abstracting Journal for Physics*

It is recommended that:

10–1. Consideration be given to the proposal for the publication, under the auspices of a single internationally controlled organization, of a single international general abstracting journal for physics, both pure and applied, including astrophysics and the geophysical sciences, and for such branches of engineering as it may be appropriate to include;

10–2. A Committee composed of representatives of the organizations responsible for the existing general abstracting services in this field, and of the interested international scientific Unions be convened to carry this proposal into effect, if it deems it desirable, by such means as giving existing abstracting journals a more international character;

10–3. This committee give attention to the proposals that the abstracts presented in the journal be mixed, some in English, others in French; and that it be in sections which might be published separately, while leaving to the appropriate time the definition of these sections and of the frontier zones for which only selected abstracts would be published.

This definite proposal for the establishment of a single *International Journal of Physics Abstracting,* preferably bilingual (English and French being both used in a single edition) had its origin in the fact that, at the time, international cooperation seemed to have in physics abstracting a simpler and more promising starting point than anywhere else in the vast field of scientific documentation. Only two journals, one printed in France, one printed in Great Britain, existed at that time to cover the whole field of the physical sciences and shared between them the international clientele of professional physicists. The possibility of an agreement between them to merge into a single international publication had been enthusiastically considered by the subcommittee which drafted the recommendation. The advantages of such an agreement—saving of time, money, and work; possible increase of means and efficacity—were clearly perceived by all members of this subcommittee. The possible disadvantage of bilingualism seemed a cheap and acceptable mutual concession, a step on the main road to international goodwill. Indeed, when the recommendation came up for discussion before the General Assembly of the International Conference, it met with general approval, and the objections of a few experienced delegates who advised that a more complete study of the problem be made before finally deciding on such a sharply outlined policy had little influence on the final wording of the recommendation, which was passed and published in the Final Act in the mist of the happy atmosphere which always surrounds the voting of counsels of perfection. The proceeding closed in the afternoon of June 25, 1949.

The International Union of Pure and Applied Physics had taken advantage of the Conference and the related presence in Paris of many scientific personalities to organize several meetings among physicists. At one of them, on June 26, 1949, the follow-up of the three recommendations examined in Sec. 1 was discussed, and it was suggested that a Joint Commission should be formed by the International Council of Scientific Unions to discuss this problem. The proposed Commission was to be composed of members representing physics and closely allied sciences and techniques; Dr. Ronald Fraser, liaison officer between ICSU and Unesco, supported this proposal. The Unions federated by ICSU were consulted and, at its Copenhagen meeting of September 1949, the Executive Committee of ICSU was able to accept this Joint Commission, which met for the first time in Paris on December 20, 1949, at the same time as a Réunion des Utilisateurs d'Analyses de Documents de Physique, organized by Unesco itself.

Professor P.Fleury, General Secretary of IUPAP, opened the meeting. After the appointment of Dr. P.Bourgeois as Chairman and Professor G.-A. Boutry as Secretary of the Joint Commission, ways and means to follow and develop the recommendations of the Unesco conference were immediately discussed. It was soon decided that the editors of the principal reviews and journals publishing original articles in physics throughout the world should be sent letters by delegates of the Joint Commission bringing to their knowledge recommendation 11 and requesting them to state in their answers:

whether the journal agreed to publish synopses of all original articles;

whether the journal agreed to publish synopses in English or in French;

whether the editor of the journal accepted responsibility of examining the synopses to make sure that they were drafted in accordance with the recommendations of the Royal Society, that they gave a correct idea of the contents of the article and were of a reasonable length.

At the time of the meeting, indeed, such an inquiry had already started in France and in Italy and Professor Boutry and Professor Perucca were able to indicate the very encouraging results which had already been obtained.

As soon as recommendations 6 and 10, however, came before the meeting, the members present realized with concern that a fact of a fundamental nature had been overlooked: no abstracting journal of any kind was represented as such on the Joint Committee; indeed the representation of abstracting journals for physics had not been in the foreground at the Unesco International Conference itself. There, mainly, the points of view of the users of abstracts, librarians, documentation services, research workers, and professors had been set forth and formulated. Clearly, recommendations 6 and 10 could not evolve in the direction of practical measures in the absence of the agencies which specialized in the making and the distribution of abstracts. The Joint Committee was unanimous in its appreciation of the fine work done both by *Science Abstracts* and the *Bulletin Analytique du Centre National de la Recherche Scientifique.* It was felt that the necessary improvements and enlargements could be discussed only in the presence of persons representing the interests and the spirit of these journals.

It was thus decided that a meeting should be organized, as soon as possible, between representatives of the Joint Committee and representatives of the two abstracting journals covering the field of physics. After having secured the assent of the Director of the *Bulletin Analytique,* the Secretary of the Joint Committee visited in London the Secretary of the Royal Society and the Institution of Electrical Engineers (responsible for *Science Abstracts*) and met there with the kindest acceptance and full approval of the initiative taken by the Joint Committee. All prior meetings having taken place in Paris, it was decided that the coming conference of “Editors and Users of Abstracts” would take place in London. After the usual consultations, the date was fixed at September 26, 1950.

The meeting was attended by:

Four representatives of the Joint Committee: Dr. P.Bourgeois, Dr. J.H. Awbery, Dean E.Hutchisson, Prof. G.-A.Boutry.

Four representatives of the abstracting journals: Dr. S.Whitehead and Dr. A.C.Menzies for *Science Abstracts;* Prof. J.Wyart and Dr. G.Kersaint for *Bulletin Analytique*.

Three observers: Dr. B.M.Crowther *(Science Abstracts),* Dr. R.Fraser (ICSU), Dr. J.B.Reid (Unesco).

The proceedings lasted for two days, during which a vital change took place in the spirit in which the whole problem had been considered before. At the opening of the discussions, this had been theoretical and didactic; when the meeting closed, it had become constructive and realistic. The goodwill and sincerity of all present, the keen realization of “users” of the excellent work which had already been done by the journals, and the generous way in which the “editors” agreed that cooperation between them could and would improve international services worked this change. Resolution 10 of the Unesco Conference was discussed thoroughly, and the possibilities of bilingualism were assessed. Most delegates, residents of European countries, including representatives of the journals, were in favour or accepted the publication in an abstracting journal of texts written either in French or English; this would enable the staffs of the two abstracting journals to divide the work between them, the

British Office concentrating mainly on periodicals printed in Great Britain and the U.S.A., and the French office dealing chiefly with publications appearing in Continental Europe. It was felt that this would result in an important saving of time and money. However, unanimity was not achieved on this point; it was explained at the meeting that the great majority of United States scientists would be strongly opposed to such a plan, and since everybody agreed that this fact could not be neglected, it was concluded that the publication of such a bilingual journal was to be conceived perhaps as an ultimate goal which could not be reached within a measurable time. Cooperation between the two journals and between ICSU and the Joint Committee must therefore proceed along other lines, which were immediately discussed. At the close of the proceedings, the main ideas which were to govern the future attitude of all persons concerned had been sifted and agreed upon. They were:

(a) That all leading periodicals publishing original papers in physics should undertake to publish authors' summaries prepared according to rules set forth by the Royal Society and the Unesco Conference of 1949.

(b) That these summaries, irrespective of the language in which the original papers were written, should be printed either in English or French.

(c) That some responsible international organization should ask the publishers of journals, whenever advisable, for special subscription rates.

(d) That attempts should also be made to obtain from the editors of important journals corrected page proofs of their issues or at least to have clippings of the synopses sent to the offices of *Science Abstracts* and of the *Bulletin Analytique* at the time of going into print, in order to speed up the editing and the printing of abstracting journals.

(e) That the editors of the two abstracting journals in physics were to cooperate thenceforward and remain in close contact, to speed up and increase the efficacity of their work, by whatever means they thought best; and that some responsible permanent international organization should be able to help them, should they require such assistance.

At this turning point in our story, let us pause for a moment to consider what progress had been made and in which direction. Briefly, the spirit of the whole undertaking had been changed. In 1949, a gathering of physicists explained to editors of abstracting journals how they should go about their business, or so these editors may legitimately have thought. A year later, almost the same persons addressing the same editors, were asking: “What can we do to help you increase the efficiency of your journals?” In 1949, it was thought that a single international abstracting journal should be founded in the near future and it was hoped that the existing reviews would be content to merge into this new undertaking. In 1950, it was recommended that two independent journals be severally helped in their work by an international organization and the impending rebirth of the German *Physikalische Berichte* already made it clear that this union might expand to include other cooperating members.

There is no contradiction in this change, only recognition that parameters other than the few considered in 1949 governed the problem; also that in international matters enthusiasm is not a factor of success, whilst humility and regard for the achievements, mental attitude, and even prejudices of others are. What was realized by all concerned was that physics abstracting was an important asset in the scientific world; that it could be improved and should be. To reach such a goal, all concerned had put aside dogmatisms of every shape and hue. After these hindrances were removed, problems in cooperation appeared simple.

In the next meeting of the Joint Committee, which was held in Paris, July 3 and 4, 1951, the logical consequences of the above resolution were considered and adopted. Meanwhile, Unesco and ICSU had been able to study the new position so that:

The Joint Committee on Physics Abstracting took note that the Bureau of ICSU accepts an International Abstracting Service as a normal permanent activity of the council. After full discussion of the implications of this decision, the Committee decided unanimously to ask the Council of ICSU to dissolve the present Joint Committee and to constitute a Board for the International Abstracting Service. On this Board, which should be small, would be represented

a) the International Council of Scientific Unions;
b) the Abstracting Journals admitted as members (initially *Science Abstracts* and the *Bulletin Analytique*)

A General Secretary should be designated by ICSU.

This suggestion was studied by Unesco and by the International Council of Scientific Unions. On May 16, 1952, in London, the final step was taken. An Abstracting Board was to be established, maintained by funds originating from Unesco and ICSU; its activity was to be controlled by ICSU, inasmuch as the General Secretary of this Council would be a permanent member of the Board, but it was to have a legal personality of its own. It would formally begin operating on June 1, 1952, and was to be incorporated in Belgium as soon as its rules and by-laws could be drafted and accepted by its first General Assembly. The Unesco Department of Exact and Natural Sciences (Professor P.Auger, Director) approved of the foundation and promised financial aid in the form of a yearly contract. Pending the first General Assembly, the Board was to be controlled by an acting Committee of which Dr. Verner W.Clapp was appointed Chairman.

The first General Assembly of the Board was held in Strasbourg, France, on July 6 and 7, 1955; all the Unions who had participated in the Joint Committee on Physics Abstracting had been asked to send observers, and the International Union of Pure and Applied Chemistry was represented by its Treasurer, Dr. Leslie Lampitt, who took an active part in the proceedings. Rules and by-laws were established and an Executive Committee was appointed whose officers and members (1953–1956) were:

Dr. P.Bourgeois (Brussels, IUA) President,

Dean E.Hutchisson (IUPAP). Members:

Prof. A.V.Hill (General Secretary of ICSU),

Prof. J.Wyart (representing abstracting journals). Secretary: Prof. G.-A.Boutry.

An office was secured in Paris; the President and Colonel Herbays applied for incorporation in Belgium; this was granted by an Arrêté Royal dated November 3, 1955. The Executive Committee has met four times, once in Brussels (July 30–31, 1954), once in Zurich (July 22–24, 1955); once in Chamonix (July 31–August 1, 1956), and in Stratford-upon-Avon (July 1957). Editors' meetings, in which the representatives of the member journals gather together with representatives of the Executive Committee are also held as often as required (London, March 30, 1954, July 30, 1954; Heidelberg, July 28–29, 1955).[1]

2. WHAT THE ICSU ABSTRACTING BOARD HAS DONE FOR PHYSICS ABSTRACTING AND FOR ITS MEMBER JOURNALS

AUTHORS' SUMMARIES

With the exception of a few publications printed in the United Kingdom of Great Britain and Northern Ireland, *all periodicals carrying original articles on physics and published in:* U.S.A., United Kingdom, Canada, Australia, Netherlands, Belgium, France, and Italy have agreed to publish authors' summaries written according to the rules approved by the Royal Society and by Unesco.

In each case, the Board has communicated with the editor responsible, who has agreed to see that the summaries prepared by the authors sum up correctly and clearly the substance of their papers. In many instances this supervision has

[1] In June 1953, the German *Physikalische Berichte* applied for membership on ICSU Abstracting Board and was admitted as the third member journal for physics under rule 3 of the Statutes (By-Laws) on November 1, 1955.

been so thorough that the member journals print these summaries *verbatim,* thereby saving the delay of abstract editing.

All summaries appearing in the countries listed above are printed in French or in English, or in both. The same results are obtained in Scandinavia and Germany. Operation has also been extended recently to the four main Japanese journals of physics which during the last few years have undertaken to print authors' summaries only in Japanese. It will be extended, in the future, to other European and overseas countries. The situation regarding authors' summaries must indeed be reviewed periodically as it tends spontaneously to deteriorate; from time to time, a few journals have to be reminded by the Board which calls their attention to the advertising value, for the journals themselves, of publishing summaries which are really shipshape. Only once in the history of the Board did an editor decline all responsibility about the matter.

PROOFS, CLIPPINGS, MICROFILMS

Fifty-five journals printed in the United Kingdom, the U.S.A., Netherlands, France, Germany, Italy, Sweden, Japan, and the USSR send the Board or its associates, generally by air mail, either complete page proofs of their issues or (in a few cases) clippings of authors' summaries. When only one copy is received, this is either microfilmed or photocopied in order that all member journals may be served. Most of this work is not centralized at the offices of the Board; to reduce expenses and delays, the editors share it and exchange photocopies and, in some cases, finished abstracts.

From time to time, the situation has to be reviewed for the reasons already stated in the preceding paragraphs. At the request of the member journals, new periodicals are sometimes asked to participate in this service; the number of these was 8 in 1954, 9 in 1955 and in 1956.

RUSSIAN LITERATURE IN PHYSICS

In 1954, the ICSU Abstracting Board, through its Chairman, got into contact with the Director of the Institute of Documentation of the USSR Academy of Sciences, Professor D.Panov. It was agreed on each side to exchange proofs of the following journals, for purposes of documentation:

Russian Journals

Zhurnal Eksperimentalnoi i Teoretischeskoi Fiziki (Journal of Experimental and Theoretical Physics)

Zhurnal Teknicheskoi Fiziki (Journal of Technical Physics)

Doklady Akademii Nauk SSSR (Communications of the USSR Academy of Sciences)

Izvestia Akademii Nauk SSSR (News of the USSR Academy of Sciences)

Uspekhi Fizicheskaya Nauk (Progress in the Physical Sciences)

Prikladnaya Matematika i Mekhanika (Applied Mathematics and Mechanics)

American Journals

American Journal of Physics

The Journal of Chemical Physics

The Journal of the Acoustical Society of America

The Journal of Applied Physics

Journal of the Optical Society of America

The Physical Review

The Review of Scientific Instruments

Reviews of Modern Physics

This exchange started in February 1955. The proofs arrived in Paris, where the Russian literature was microfilmed and sent on in this form to the editors of the member journals. This attempt to improve abstracting on both sides was hampered, at its beginning, by mailing and reproduction delays. The situation improved gradually; today, however, direct exchanges between members in most cases have been substituted for the former procedure.

Upon the request of the American Institute of Physics, which contemplated publishing translations of one or more of the important Russian journals of physics, the Board has instituted an inquiry in twelve European countries to assess the feasibility of this scheme. The results of this inquiry were published in a pamphlet (April 15, 1955) “A Study of the Feasibility of a Comprehensive Russian to English Translating Service in the Field of Physics” (Document IAB 116R/55) in which the statistics of 571 questionnaires are given. These results were communicated to the American Institute of Physics, which then announced the publication, to start in November 1955, of a full English translation of the Russian *Journal of Theoretical and Experimental Physics*. This news was communicated by the Board to the USSR Institute of Documentation, which expressed satisfaction (Chamonix, 1956). The success of this enterprise of the AIP and its extension to other major Russian journals is now a fact.

Acting jointly with the Publication Committee of the International Union of Pure and Applied Physics, the Board also helped to organize the work of publishing review articles describing the activities and progress of the principal schools of research in physics of the USSR and other Slavonic-speaking countries; in three supplementary numbers *Il Nuovo cimento* has already published such articles, which are written in English or in French; all carry extensive and classified bibliographies of Russian papers.

CLASSIFICATION

The UDC classification in physics is now undergoing revision. This move, which had been urged by the Board since 1952, has originated in the Fédération Internationale de Documentation. The Committee on Publications of the International Union of Pure and Applied Physics and the Board have joined forces to cooperate on this revision, and both are represented on the FID special committee by the editor of one of the member journals. It has already been decided that, as soon as this revision is finished, the new classification will be used uniformly by all member journals of the Board.

NONPERIODICAL PUBLICATIONS

As is well known, research workers and libraries receive free of charge a great quantity of semiperiodical or nonperiodical literature which they do not always require, while they may often be unable to learn what is contained in some report or account of a Congress in which they are interested. Two moves have been made to facilitate abstracting work in this field:

(a) The editor of each member journal is now dealing with nonperiodical publications of a group of countries and directs the attention of his fellow-editors to these publications which he considers of interest to physicists.

(b) The Board has asked the International Union of Pure and Applied Physics, which in almost all cases makes grants to the organizers of congresses and colloquia on physical subjects, to impress upon organizers the importance of providing proper means of publication of the proceedings of these meetings, preferably in recognized journals, and, if they issue special reports of proceedings, to send these publications to the Board to ensure thorough and competent abstracting.

CHEMISTRY

The abstracting of physical literature is not now the only activity of ICSU Abstracting Board. As early as the end of 1950, an active and shrewd Executive Committee of the International Union of Pure and Applied Chemistry had expressed the opinion that the “Joint Committee” then preparing the birth of the ICSU Abstracting Board, should concern itself with the abstracting of both chemical and physical literature. This wish was endorsed by ICSU in a Wash ington meeting of October 1951. Somewhat later, a proviso was added that work on chemical documentation should not be attempted before sufficient progress had been made in the field of physics. Acting upon this suggestion, it was only in 1954 that the Board, in its yearly report to ICSU, intimated that it felt ready to extend the scope of its activities. ICSU authorized this extension in its Naples meeting of 5–7 October 1954. The International Union of Pure and Applied Chemistry upon being notified of this decision, recommended the candidatures of *Chemical Abstracts* and the *Bulletin Analytique* as member journals for chemical abstracting in English and French respectively. These journals were unanimously elected to the Board. In 1956, the German abstracting journal *Chemisches Zentralblatt* also applied for membership and was elected unanimously at the 1957 meeting. Dr. R.Morf sits on the Board as representative of IUPAC. Work has now begun, and the editors of a select number of important chemical journals are already being communicated with, with a view to arranging page proof exchange of their publications among members for chemistry. The report to the General Assembly of 1956 tabulated the first results, which are being steadily increased on the same basis already used in physics. However, because the world of chemical documentation is so much larger than that of physics, personal contacts are proportionally harder to establish. The efficiency of the Board would therefore have shown a tendency to decrease if no new policies had been adopted. Thus it came about that the Board appointed, in a few countries of major importance to chemical documentation, correspondents who are young scientists already sufficiently known to have a sympathetic audience in the chemists and editors of their nations. So far, correspondents have been established in Japan, Germany and Switzerland, Scandinavian countries, and Great Britain. The Board is contemplating extending this procedure to Latin America, the United States, and India.

A special case is that of the USSR. For the present, the Russian Institute of Scientific Information is acting as the Board's correspondent for Slavonic countries. By these means a very active exchange of periodicals has been instituted and the USSR Institute of Scientific Information has procured for the Board member journals proofs of all the chemical periodicals published under the sponsorship of the Russian Academy of Sciences.

THE FUTURE: TRENDS AND PROPOSALS

The aims and motives of the ICSU Abstracting Board should now be clear to the readers of the present paper. To go on improving the quality and the speed of availability of scientific information, the ICSU Abstracting Board must continue to promote an ever closer voluntary international cooperation among all parties concerned: authors of papers, editors and publishers of journals, editors and publishers of abstracting journals, librarians, and documentalists. It must secure the general approval of a code of ethics applicable to scientific information, its manufacture, and its distribution. The experience of the past years has shown that this can and should be achieved without taking any measures having a coercive character. It is therefore proposed to ask for the support of all interested international organizations and unions concerned for a campaign to be launched by the Board in the near future. The following proposals would be stressed.

1. Authors of papers should never submit a paper in manuscript to an editor without forwarding at the same time an abstract drafted by themselves in two versions, at least one of these being written in either English or French. These abstracts should be composed in accordance with the Guide for the Preparation and Publication of Synopses (document UNESCO N.S. 5I D 10 aA/05 XI.5I).
2. Editors of journals should voluntarily agree to see that the papers published in the journals be preceded by an author's abstract. The editor or his delegate should personally verify that these abstracts are drafted according to the Guide for the Preparation and Publication of Synopses.
3. Editors of journals should voluntarily agree in favour of IAB's members to waive copyright in the case of these authors' abstracts only; that is to say, they should explicitly permit free reproduction by any member journal of IAB of any of these abstracts. Editors should realize that this facility will greatly favour the circulation and advertising of their publications.
4. At the time of going to press, editors of journals should agree to send, by a quasi-automatic procedure, clippings of all the authors' abstracts in the issue either to the General Secretariat of the ICSU Abstracting Board or to the member journals of their choice. Failing this, and if more convenient, clippings could be replaced by the mailing of advance copies.
5. National committees of science and national scientific societies should be persuaded to help in tracking and listing serial or nonperiodical publications in their countries. A quarterly list of these publications should be made for each country and sent to the General Secretariat of ICSU Abstracting Board which would circulate these lists to all its member journals.
6. National committees and national scientific societies should agree to discourage excessive printing of nonperiodical publications; they should insist that proceedings of congresses, symposia, etc., held in their country be published as

far as possible in internationally recognized journals. Failing this, they should see that the publication of the proceedings of congresses, etc., are included in their quarterly list of nonperiodical publications.

It is hoped that all the participants to the Washington Conference on Scientific Information will agree that these proposals will not in any way restrict the freedom of authors and publishers of scientific news and scientific information, though their general acceptance should greatly speed up and facilitate the diffusion of high-quality scientific original literature throughout our world of today.

Creation of an International Center of Scientific Information

PAUL BOQUET

Because of the rapid progress of human knowledge, we produce a considerable quantity of documents. The increasing number of scientific books, the increase of specialized periodicals, and the accelerated rate at which international conferences and colloquies are being organized render the work of those who devote themselves to scientific research, an essential part of which consists in retrieving information indispensable to the progress of their work, more difficult every day. Some progress has been made in the field of documentation since the last century, but it should be recognized that the means which we possess no longer meet the requirements of the new disciplines.

The reasons for adapting methods of abstracting, classification, and dissemination of scientific documents to these requirements appear on reading the proceedings of discussions by the most eminent specialists. Certain of their discussions have already borne fruit. The effort of simplification and rationalization accomplished in the area of terminology, abbreviations, and scientific symbols, in the improvement of methods of classification and selection of information, the measures suggested in the conclusions of the International Conference on science abstracting, the agreement established among the editors of several large abstracting journals, the development of translation services, in short, the work of world-wide cooperation accomplished by the International Council of Scientific Unions—all are evidence.

Nevertheless, in spite of the increase of information centers in all countries (apart from the Centre National directed by Professor Wyart, there are more than three hundred in France), our methods of documentation remain imperfect, and every day the necessity to improve them by establishing close cooperation among the specialized services is felt more and more.

To meet this need, would it not be wise to unite our effort by setting up an international center as coordinator of all documentation services?

Although it is difficult initially to delimit with precision the immense field of

PAUL BOQUET Institut Pasteur, Paris.

activity of such an institution nevertheless we shall try to state briefly its main functions and to define its relations with other information services. In short, without minimizing the difficulty of the problem we are raising, we shall examine some of the indispensable conditions for the foundation and development of this institution.

The facts drawn by researchers out of the documents at their disposal are of various kinds. Some are analytic. They include indicative cards, abstracts of articles, of theses, and of scientific works, indexes, tables of contents, lists of periodicals, texts translated from one language into another, and finally, journals devoted to the publication of summaries. Others are synthetic. These include synoptic accounts, monographs, general reviews followed by extended bibliographies, and reports in extenso of conferences or symposia dedicated to the study of a particular problem. Still others are of a technical character. They are concerned with scientific apparatus and patents.

The activity of the international center would consist not only of *gathering, reproducing,* and *disseminating* information but also of *close cooperation with all documentation services*. Each service or each national or local center sending to the international institution information published in its own country, would get in exchange documents of foreign origin translated into one of the officially adopted languages. Thus the abstracts of the scientific texts, published in each country (abstracts written by professional abstractors, author abstracts, homotopic abstracts, or indicative abstracts, according to the circumstances), collected, classified, and translated into several languages through the efforts of the international center, would be distributed among the editing committees of all abstracting journals.

Technical information, especially that concerning scientific apparatus, would be collected and then distributed by a similar procedure.

Because of the quality of the information contained in the monographs and periodicals devoted to the discussion of current problems or to the publication of reviews, the international center will have the responsibility of publishing versions of these publications in *several languages* in order to encourage their propagation and diffusion.

In the same line of thought, *conferences* and *symposia* organized through this institution, would, at regular intervals, bring together the specialists of the most diverse fields of scientific disciplines.

The proceedings of these meetings would be published by the international center. This organization, in short, would meet the desire of all scientists if it contributed to the establishment of links between scientific research centers, no matter how distant they might be, not only by asking for exchanges of documents, but also by placing at the disposal of the scientific public a directory of these services and lists of specialists. In this regard we will mention as an example the directory of laboratories recently published in France, under the direction of M.Bayen and H.Weiss, by the national office of French universities and schools.

The accumulation of materials by the international center would make indispensable the use of improved methods of classification, selection, and reproduction of documents. We shall not attempt to enumerate these methods here, or to summarize the discussions which they raised, but, in order to explain our idea, we shall quote the opinion of H.Coblans who writes:

Even if not immediately profitable or feasible, some form of automation in the long run is one of the most likely remedies for the documentation chaos already engulfing us (*Unesco Bulletin for Libraries 11,* 11 (July 1957).

We do not doubt that some have certain reservations concerning the value of electronic devices, but we hope that the continuing improvement in devices will contribute to the spread of their use for registration, classification, selection, and translation of scientific information, and will thus place all sources of cybernetics at the disposal of the international center.

The purpose of the bibliographical services within this institution would be to perfect its multiple functions in close cooperation with similar services of other centers and of those of industries whose activities are related to it. The same services would be in charge of *standardizing bibliographical methods* and *improving the procedures for the dissemination of information* in solving the problem raised ten years ago by Professor Bernal about a possible eventual reform of the structure of scientific periodicals and the method of their distribution.

Thus guided by its own bibliographical services, enriched by documents arriving from all countries and invigorated by continuous exchanges the international center of information would be called to play an essential role in the development of scientific research.

HOW TO ESTABLISH SUCH A CENTER, IN THE ACTUAL CIRCUMSTANCES

It is difficult to imagine the creation of an entirely new organization which would impose on each nation a heavy financial burden, but we may justifiably imagine that the institutions founded with the purpose of establishing international cooperation in the field of scientific culture would respond to the appeal addressed to them. I will cite Unesco, whose Director-General at its General Conference at New Delhi saw it entrusted with the task of encouraging the organizations that were founded for the purpose of developing and improving documentation methods.

In a recent report Dr. Bentley Glass suggested that this institution, the Inter national Council of Scientific Unions, the International Union of Biological Sciences, the International Organization of Medical Sciences, and the individual biological societies should be invited to give financial help to the abstracting journals in order to allow them to participate in the vast plan of cooperation that he presented. The cooperation of these foundations, that of the World Health Organization and various scientific societies would be necessary for the creation of the international information center, but it is doubtful whether these institutions, solicited from all sides, would be able to furnish sufficient material aid. For that reason it is necessary to consider appealing to the resources of organizations whose prosperity is closely tied to the development of scientific research. A contribution from industry would place at the disposal of the information center financial means commensurate with the work to be accomplished.

On the other hand, it would be desirable to persuade countries to abolish customs duties and lower postal rates for anything related to the exchange of scientific documents.

One might object that those governments with the greatest financial resources might exercise an overwhelming influence within the international organization. It must be admitted, however, that the constitution of this institution would have to assure its moral independence.

We may equally object that the collecting of documents by an omnipotent organization and the standardization of scientific information methods could, after a short time, bring about the disappearance of ancient institutions. Certainly the development of the international center would render obsolete manual methods of documentation, but we may hope that the innovators would avoid any destructive revolutions by obtaining the collaboration of all existing organizations which will adapt their activities and techniques to the new system of exchange. In order to obtain coherent teamwork, it would be useful to define and coordinate the activities of these organizations. Thus learned societies, scientific institutions, universities, certain government departments, and industrial research centers would be called upon to cooperate, in each country, in order to render effective these innovations, to assure a rational division of work among the information services, and to establish an efficient control over this work.

One might finally allege that a center with the purpose of establishing international cooperation in the field of scientific information would see its role rapidly restricted by strategic reasons imposed by the circumstances. There can be no doubt that today frontiers restrict several fields of science and interrupt the exchange of documents, but other fields, e.g., medicine, biology, and the related sciences are widely open and lend themselves to the fundamental experi ment that the creation of the center would be. Thus the rhythm of the researches carried out for peaceful purposes would be increased to the benefit of humanity.

PROPOSITIONS

It would be desirable through such institutions as Unesco and ICSU to carry out an inquiry on the advisability of creating an international center devoted to enriching the collective patrimony by the exchange of documents between all countries.

The goal of this center would be:

1. To assure close cooperation between all documentation centers (national and regional centers, centers operating under state control, as well as private centers).
2. To collect, classify, select, preserve, translate, and reproduce scientific information by modern methods.
3. To facilitate the dissemination of periodicals and selected monographs by publishing them in the officially adopted languages.
4. To organize symposia and conferences and to publish their reports.
5. To publish at regular intervals indexes, tables of contents, and journals devoted to summaries of periodicals.
6. To carry out bibliographical work for scientific institutions, laboratories, and offices of various industries.
7. To organize research services for unifying and improving documentation methods.
8. To aid relations between all scientists by publishing a directory of research centers and lists of specialists.

In order to carry out this program the financial participation of official institutions (international institutions, learned societies, various foundations), and even private groups (publishers, financial companies, and industrial groups) might be solicited.

An International Institute for Scientific Information

WALDO CHAMBERLIN

We are living in a society in which scientists "will largely determine the fate of mankind,"[1] a society in which "there is every reason to believe that the pace of change will be much more rapid than during the previous period in our technical development,"[2] and one in which there are not enough scientists. In such a situation, the storage and retrieval of scientific information is a vital and urgent problem.

THE PROBLEM

The problem can be stated quite simply: more scientific information has been produced than can be stored and retrieved, and the body of that information is said to be doubling every ten years. Thoughtful men have long been aware of the resulting dilemma, and as early as 1936 Ortega y Gasset clearly stated the issue in an article entitled, "Man Must Tame the Book."[3] In 1945 Vannevar Bush pointed out

The difficulty seems to be, not so much that we publish unduly in view of the extent and variety of present-day interest, but rather that publication has been extended far beyond our present ability to make real use of the record. The summation of human experience is being expanded at a prodigious rate, and the means we use for threading through the consequent maze to the momentarily important item is the same as was used in the days of square rigged ships.[4]

WALDO CHAMBERLIN Professor of Government, Graduate School of Arts and Science, New York University, New York.

[1] BUELL G.GALLAGHER, *New York Times,* November 15, 1957.

[2] Report of President's Committee on Scientists and Engineers, November 30, 1957, as reported in *New York Times,* December 1, 1957.

[3] The *Wilson Bulletin,* January 1936.

[4] "As We May Think," *Atlantic Monthly,* July 1945, p. 102. This was the article in which Bush suggested, with tongue in cheek, perhaps, his famous "Memex"; a desk in which a scholar could keep all his records, notes, clippings, etc., for a lifetime yet have them instantly available through the use of electronic devices. See also his *Science, the Endless Frontier,* U.S. Office of Scientific Research and Development, Washington, 1945.

The need for methods of storage and retrieval that are better than those used in the "days of square rigged ships" stems from two facts; the first being that we do not have, and are not likely to have, enough scientists to meet the foreseeable demand for many years to come and, second, that the scientists that we do have waste an immense amount of time in determining what has already been done in any area of research, a process sometimes called "literature search." The scientist makes this search in order to avoid duplicating work already done elsewhere, and because he must know what others have done, or found out, that will assist him in his own project.[5]

WHAT INFORMATION DO SCIENTISTS NEED?

Scientists seem to be in general agreement that there are six major problems of storage and retrieval that apply to all branches of science in varying degree:

1. Lack of *comprehensive* means for retrieving information, particularly in peripheral fields.
2. Lack of accessibility to many publications.
3. Lack of translations of much of the material appearing in "exotic" languages, such as Russian and Japanese.
4. Delay in publication of abstracts.
5. Uncertainty that abstract coverage includes important articles and reports in little known periodicals, or from "remote" scientific meetings.
6. Duplication of effort and frustration in literature search.

WHAT HAS BEEN DONE ABOUT THE PROBLEM?

The amount of time that scientists spend on literature search is not known. We do know that the process is so involved and complicated that some industrial concerns have established technical information divisions because "digging out information has become so time-consuming that men are discouraged from making the effort"—this "at a time when technical manpower is at a premium."[6]

Studies of the problem have not been lacking, and evidence that there are at

[5] John P.Alden of Food Machinery & Chemical Company's technical center has said that there are two ways to solve a technical problem: the laboratory and the library. He notes that library research is probably cheaper than laboratory research, and calls attention to the fact that companies that report primary dependence on library sources for information for management decisions are in the first 30% of companies with increased earnings. *Chemical and Engineering News,* Sept. 23, 1957, p. 115.

[6] One of these is the Esso Research and Engineering Company which has established a Technical Information Division that is described in an article in *Petroleum Weekly,* March 1, 1957, entitled "System of the Future Gets Information to the Technical Man Who Needs It." The new Division is also described in a paper presented before the American Chemical Society's Division of Chemical Literature, by three officers of the company, W.C.Asbury, E.Duer Reeves, and W.T.Knox.

least partial answers available is provided by the program of the International Conference on Scientific Information. There have been so many conferences, symposia, colloquia, round tables, workshops, and seminars devoted to the subject during the past ten years that there is a temptation to apply to the problem Mark Twain's famous remark about the weather, "Everybody talks about it but nobody does anything about it." In 1949 the National Research Council rejected a plan for an international coordinating office for science abstracting because it was "quite without prospect of the necessary financial support, and on the further grounds that the first thing for the United States to do was to put its own house in order." Insofar as financial support was concerned, the International Conference on Science Abstracting and Indexing, convened by Unesco in 1949, was of the same mind, but did recommend some "palliative measures."[7]

While these "palliative measures" have been under way, the Russians in 1952 established an All-Union Institute of Scientific Information, in an attempt to provide a complete, coordinated system of documentation for their scientists.[8] Regardless of whether or not this Institute has accomplished all that was expected, it is an attempt to tackle the problem of storage and retrieval on a coordinated basis.

In spite of efforts by individuals and groups that have tried to do something about the problem of storage and retrieval, the conclusion is unescapable that what has been accomplished to date has been a whittling on pieces of an immense problem.

WHAT SHOULD BE DONE ABOUT THE PROBLEM?

The situation has now reached a point where it is, or should be, clear that the time for palliative measures has passed. The President of the United States has said that in this country there has "been a tragic failure to secure the benefits that would flow from mutual sharing of appropriate scientific information."[9] The political and scientific atmosphere seems to be ripe for a great step forward in the development of a huge program of basic research—a necessary part of which would be the creation of adequate tools for storage and retrieval of scientific information. The problem is so large and so complicated that it probably cannot be solved without major financial assistance from govern

[7] Verner W.Clapp, "Indexing and Abstracting: Recent Past and Lines of Future Development," in *College and Research Libraries,* July 1950, p. 203.

[8] For accounts of the Institute, see Rita G.Piepina, "Scientific Documentation in the Soviet Union," in *American Documentation,* vol. V, pp. 71–79, January 1954; "Scientific Information in the U.S.S.R." in *Science,* vol. 124, No. 3223, 5 October 1956; and Harry Schwartz, "Soviet Science at Work," in *New York Times,* November 17, 1957.

[9] *New York Times,* December 8, 1957.

ment. The difficulty of doing anything truly effective lies in the fact that long-range planning with financial support is necessary and, in the United States at least, long-range planning and financial support by Government are still feared as being "creeping socialism," or worse. The pressure of events, however, may be such that this fear can be overcome. Perhaps the establishment of the All-Union Institute for Scientific Information in the USSR may have been such an event. The issue for the United States and the countries of the West is whether or not they can afford to continue their uncoordinated, unrelated, inefficient, and wasteful methods for storage and retrieval.

In considering the methods for solving the problem, scientists seem to divide into three groups. One group contends that the emphasis in the approach to storage and retrieval should continue to be on study, rather than on action to solve the problem. A second group believes that we should not jump too rapidly in any direction until more thought has been given to the problem, but that an increase in cooperative action is needed now. A third group argues that we already have sufficient knowledge of the nature of the problem to make it possible to take a major step forward, even though research on the nature of the problem must continue. The first group believes that considerable progress on storage and retrieval has been made and that any radical move might endanger what has been accomplished. Chemistry and medicine are cited as two branches of science in which there have been great accomplishments through cooperation but without coordination.[10]

The second group acknowledges the accomplishments made thus far, but believes that the situation is so acute that a much greater cooperative effort is now necessary. This group sees the great work in chemical abstracting, for example, as a splendid achievement, but it believes that broader coverage and more rapid abstracting are necessary, and that they can be achieved through improved and increased cooperation of all interested parties. This group also believes that the development of bibliographical controls in medicine is impressive but that there are too many such controls and that some rationalization and simplification can be achieved through increased cooperation.

The third group, the one desiring coordinated action now, differs from the second primarily in its conviction that the present need is so great, and becoming greater, that cooperation without coordination will not provide an adequate solution to the problem of storage and retrieval.

Within all three groups there are individuals who believe that the most effective approach to the problem must be international in scope, because science is international, and because the material in question is already within

[10] There are said to be 450 bibliographic services in the field of medicine. See Council on Library-Resources, *First Annual Report,* Washington, 1957, p. 15.

the public domain and not affected by security regulations.[11] Any effort to solve the problem by a purely national effort would be wasteful and inefficient and, in the long run, more expensive than an international effort.

A coordinated national attack on the problem of storage and retrieval would mean, in the United States, an effort by Government, industry, universities, research institutions, professional societies and philanthropic foundations. It might be less difficult to secure adequate financial support from Government if the problem were attacked on the national level, but the international character of science would make such a program inefficient and wasteful. Scientists need to know what is going on in other countries and without a pooling of effort, each nation duplicates many of the services provided by others and, because resources are limited, certain needed services are not supplied at all. For example, there are said to be 150 abstracting services that are of interest to physicists and 450 in the field of medicine, but none in the English language for astronomy or geophysics. Further evidence of duplication of effort and serious gaps in bibliographical controls is the fact that of the articles abstracted in *Physics Abstracts,* 47 percent were also abstracted separately for *Chemical Abstracts,* and this at a time when about two-thirds of scientific articles are not abstracted at all.[12]

RECOMMENDATION FOR AN INTERNATIONAL INSTITUTE FOR SCIENTIFIC INFORMATION

The best solution to the problem of storage and retrieval will probably be found in the development of an international organization, the fundamental purpose of which would be to assist the scientists of all nations, by providing the machinery through which nations could do together those things that can be done effectively only when done together. An international institute for scientific information could provide for all scientists what the Soviets are trying to provide for theirs. The best elements of the Russian experiment should be used, but to these should be added the advantages that are inherent in competition, flexibility, and cooperation; advantages that are seldom found simultaneously in any governmental body.

To achieve such a purpose, the powers and authority of the institute would have to be clearly defined. It should have sufficient autonomy to “spit in the eye” of any government, industry, university, professional society, or founda

[11] This paper is not concerned with classified scientific information, which is already being considered as the result of the NATO meetings in December 1957. See the “Declaration of Principles,” published in the *New York Times,* December 20, 1957.

[12] Clapp, p. 198.

tion that attempted to dominate. Because of the very great services that could be rendered by an institute sufficiently autonomous to be free from domination by any vested interest, except that of international science, governments should find it to their advantage to permit such freedom, and to provide the necessary financial support. It is obvious that states that supplied the major funds would influence the character of the institute, but influence and domination can be quite different.

AUTONOMY OF THE INSTITUTE

The autonomy of the institute might be assured if its resources were diversified so that no government (or group of governments), industry, professional society, university, research institution, or foundation would be in a controlling position. The necessary diversification might be obtained by including in the formula for national contributions both money and services. Certain functions of existing national scientific bodies, both public and private, would be adapted to the needs of the institute and included in the national contribution. For example, the American Chemical Society's expenditures for *Chemical Abstracts* could be included in the United States contribution, and the sum required to publish the Soviet Union's *Abstract Journal-Chemistry* could be included in the Soviet contribution.

Services presently operated by government agencies could also be included, such as the Department of Commerce clearing house for scientific information, and the National Science Foundation program for translation of Russian scientific works.[13] Similarly, certain services now provided by private organizations could be adapted and included in the national contribution. To accomplish this adaptation, something more than cooperation is necessary. What is required is coordination by an international institute that would have sufficient resources to make cooperation advantageous to potential cooperators in all countries. In other words, a major source of the institute's authority would be the fact that it would have resources that could be used to assist national agencies to do better the jobs that they are now doing and, with the same resources, channel some of the duplication into areas of science not now being served.

LEVELS OF SCIENCE TO BE SERVED

It has already been noted that certain fields of science have better storage and retrieval tools than others, and the same holds true for the different levels of scientists. For purposes of discussion, these may be said to be: the advanced

[13] The National Science Foundation has a $107,000 translation program. *New York Times,* November 25, 1957.

scientist, the developmental scientist (college instructors and institutional researchers), the beginning scientist (the undergraduate), and the science teacher in secondary schools. The international institute should coordinate the efforts to meet the storage and retrieval needs of all four of these levels. All of them have certain common needs, all are dissatisfied with the present state of things, and all seem to agree that the proper use of the scientists that we now have, including those that are being trained,[14] requires considerable improvement in methods for storage and retrieval at all levels.

THE CHARACTER OF THE INTERNATIONAL INSTITUTE

This world-wide problem should be approached on a universal international basis, and there is an organization available that is appropriate to undertake such a task, the United Nations Educational, Scientific and Cultural Organization (Unesco). If all nations were convinced, as some appear to be, that the work of their scientists is of vital importance to the future of their countries and that the needs of their scientists for adequate facilities for storage and retrieval are acute, it should not be impossible to secure agreement upon the establishment of the International Institute within the framework of Unesco. It is generally known that international cooperation in any field is difficult to achieve, and that success is often more likely when the international body concerned has funds that can be used to make cooperation advantageous, or to turn limited cooperation into real coordination. There are many observers who believe, for example, that the United Nations Expanded Technical Assistance Program has achieved what it has, because it had the funds to make cooperation advantageous and coordination possible.

Should there be substantial disagreement that a world-wide international institute for scientific information within the framework of Unesco is not desirable, it could be established as a separate organization. If there were not general acceptance of the idea of a world-wide institute, it could be developed within the framework of the North Atlantic Treaty Organization (NATO). However, because science is completely international, a world-wide organization would not only be desirable but also would be the only likely means of providing the scientists of any nation with all the information that they need. Any regional concept would be second best and should be avoided unless it were certain that the broader approach was impossible.

---

[14] Among the many who have suggested the vital importance to the future of improvement in the quality of science teaching at the high-school level, is Fred Hoyle, a British scientist presently serving as Visiting Professor of Astronomy at California Institute of Technology. See *New York Times Magazine,* January 24, 1958.

POLICY MAKING IN THE INSTITUTE

The policies of an institute designed to provide *all* scientists, at *all* levels, with *all* the information that they *need* should be determined by scientists, subject to influence, but not control, by governments. Storage and retrieval of scientific information cannot be controlled by government, but it is a problem of such magnitude that governments must provide major financial backing. The general policies of the institute could be set by representatives of states in a body such as the General Conference of Unesco, but executive policy should be determined by scientists. This might be accomplished by the creation of an *executive board* consisting of one member from each of fifteen states, with the states being selected by the General Conference of Unesco. The determination of the individuals that should sit on the executive board should not be made by governments, but by *national scientific information commissions*. These commissions should consist of very eminent persons, predominantly scientists, in each country. In the United States, for example, the commission members might consist ex officio of the presidents of such bodies as the American Council of Learned Societies, the National Research Council, the National Academy of Sciences, the American Documentation Institute, the Association of American Universities, the American Library Association, the National Education Association, and the learned societies in each field of science. Commissions composed of individuals of this type could be relied upon to insure that no interest but that of science would dominate.

The type of guidance that the institute would need to attack the very wide range of problems involved in storage and retrieval could be provided by the national commissions. They should also be able to command sufficient national support to insure adequate action by legislatures in providing necessary funds, and to secure cooperation from governmental agencies, universities, industry, professional societies, foundations, and research institutions. Because most of the commission members would have had considerable administrative experience, they should be well qualified to select the proper type of national representative to sit on the executive board.

The primary function of the executive board would be to select a full-time *president,* and to assist him in determining the executive policies of the institute. It would be the responsibility of the national commissions to select as members of the executive board, persons who had had experience in working with, and would have personal knowledge of, the type of individual needed for the presidency of the institute.

ADMINISTRATION OF THE INSTITUTE

The president might be a scientist; he would have to be sympathetic to, and understand the needs of scientists. It would be desirable that he have experience in science, government, industry, and education. Because his task would be primarily one of securing the continuing support of governments, industry, universities, foundations, professional societies, and research institutions, he would have to be a negotiator, a salesman, and an organizer. A combination of such qualities in one man is rare, and the compensations would have to be maintained on a very high level. However, the challenge of the job itself would probably be more significant in attracting the proper type of person to the job than would large financial gain or prestige.

The structure of the institute should reflect the major areas where coordination of activities in storage and retrieval is most needed. Whatever the form of organization, it would have to be flexible, because the problems with which it would be concerned are interrelated. The major needs of scientists suggest an institute with six departments: abstracting, indexing and cataloguing, status of science, translation, library and information centers, devices. The work of each department would be related to and influence that of all others. The abstracting department would have to work closely with the indexing and cataloguing department, and both would be assisted by the translation department. All three would have to cooperate with the status of science department in the development of new approaches to the theory of knowledge, and in the preparation of compendia, reviews, surveys, and other publications designed to help the various levels of scientists keep abreast of their particular fields. These four units would aid the library department in developing better procedures, processes, and equipment for use in scientific libraries all over the world.

The devices department would assist and be assisted by all other departments in the development of new equipment for abstracting, indexing, cataloguing, publishing, and physical storage and retrieval in libraries. All departments would cooperate in developing information centers, and every department would have as a continuing function participation in the training of storage and retrieval experts.

FINANCING THE INSTITUTE

The palliative measures that are used to solve the problem of storage and retrieval of scientific information are the result, as noted above, of a general belief that there is no likelihood of securing funds commensurate with the job that needs to be done. There is agreement on the need; there is even a wide area of agreement on what should be done to meet the need. What seems to be lacking, first of all, is a willingness to state the need in terms of money required to tackle the whole problem of storage and retrieval, rather than in terms of the money required for individual bits and pieces of the problem. I am aware that fools step in where wise men fear to tread. Nevertheless, if someone will say that an international institute for scientific information that would provide adequate storage and retrieval of scientific information would cost about $283,350,000 a year, wise men may be willing to refine that estimate and, what is more important, make an effort to induce governments and the scientific community to provide the necessary funds.

Listed below is a recapitulation of expenditures for an international institute such as is envisaged in this paper.

Annual budget for an international institute for scientific information

Administration of the institute		$20,000,000
Abstracting department		
Preparation of abstracts	$18,750,000	
Transmission of abstracts	18,000,000	
Publication of abstracts	120,000,000	
Research staff, equipment, etc.	10,000,000	$166,750,000
Indexing and cataloguing department		
Indexers	3,600,000	
Research staff, equipment, etc.	1,000,000	4,600,000
Status of science department		
Journals and yearbooks	12,000,000	
Compendia and literature improvement	2,000,000	14,000,000
Translation department		
Translation	44,000,000	
Research and equipment	5,000,000	49,000,000
Devices department		
Supply of machines	5,000,000	
Development and research	5,000,000	
Supply of retrieval systems	10,000,000	20,000.000
Library department		
Library schools	2,500,000	
Information centers	1,500,000	
General assistance to libraries	5,000,000	9,000,000
Total		$283,350,000

An institute having the functions listed would not involve an additional outlay of $283,350,000 because certain existing national expenditures would be credited toward this total.

Some readers will immediately contend that $283,350,000 is an “impossible” total. The difficulty in giving an effective answer to this charge lies in the lack of any precise figures that show the cost of the present unintegrated, uncoordinated, unorganized, overlapping, and duplicating system of storage and re trieval of scientific information. However, there are some figures available in regard to the cost of science abstracting services that suggest that the above figure might be less than what is now being paid for the inadequate system under which scientists have to operate.[15]

There are 3400 known abstracting services in the world, and it is improbable that the average cost of each is less than $50,000, or a total of $172,500,000. *Chemical Abstracts* costs about $1,300,000, and we know that there are at least twenty industrial concerns in the United States alone whose expenditures for abstracting services are closer to $100,000 than to $50,000 a year. It would be surprising if the abstract services of the Soviet Institute for Scientific Information cost less than $5,000,000 a year. The expenditures of the twenty or more governmental units in the United States that do abstracting may be close to that figure. How far from the mark would it be to say that the world now spends $200,000,000 for science abstracting?

We do not know what it costs to have no abstracting services at all in the English language in fields such as astronomy and geophysics. Furthermore, we can only guess at the savings that might result, for example, from a reduction in the 450 indexing and abstracting services in the field of medicine[16] or from elimination of a portion of the double abstracting of the same material, or from a substantial reduction in the 3400 indexing and abstracting services in the world, or from an amalgamation of some of the 150 abstracting services in the field of physics.

Of the things that are known that support the case for a world-wide attack on all phases of the storage and retrieval problem, the most important are the facts that science is international and knows no boundaries and that the problem of storage and retrieval in one country is essentially the same as that in all countries—the difference is only one of degree. Therefore, the attempt of any one country to "put its house in order" is, at best, a second rate answer, because no national science can stand by itself. The development of an effective national storage and retrieval system requires absolutely that the scientific production of other countries be included. The creation of effective systems in any two countries would require that each duplicate almost completely what was done in the other. The creation of effective systems in five countries would make it necessary for each to duplicate most of what was done in the other four.

The failure of governments to remedy the situation probably lies in the inertia of politicians in respect to anything that is thought to be non-political.

---

[15] A similar hypothesis may have been in President Eisenhower's mind on November 7, 1957, when he said that "the mutual sharing of appropriate scientific information" was "one way in which, *at no cost* (italics mine), we can dramatically and quickly magnify the scientific resources at the disposal of the free world." *New York Times,* December 8, 1957.

[16] Council on Library Resources, *1st Report,* p. 15.

If scientists agree that the problem of storage and retrieval is too big to be handled properly without an assist from government, the problem must be made "political."

BRINGING THE INSTITUTE INTO BEING

If scientists expect politicians to take an interest in science, scientists will have to take an interest in politics. A first step toward making storage and retrieval political might be for a group of national scientific organizations, professional societies, and universities to make a very strong recommendation that an international institute for scientific information be established as quickly as possible. If the range of scientific interests in such a group were broad enough, the extent of their influence, if they really desired to exert it, should bring forth the necessary initial support, which would have to be both political and financial. Political support would almost surely come from politicians who are also statesmen, men who can see that what sometimes appear to be immediate issues of the moment can only be solved by long-range policies and programs. If a group of organizations such as is suggested were convinced that an international institute is needed, it is probable that educational foundations, philanthropic individuals, and industry would provide the funds necessary for the preliminary work required to bring the institute into existence.

Is it too much to suppose that a group of foundations, or even one of them, would provide the necessary funds for the preparatory work if it were clear that the scientists of the world believed that an international institute for scientific information is needed, and that it was something toward which they were willing to work?

SUMMARY OF DISCUSSION

The Chairman of the Panel opened the Area 7 session with the statement that the discussion now comes down from the pure air of mathematical analysis and theorem-building to the smoke-filled room of politics, economics, and administration. He called attention to the fact that, in addition to the nine papers brought together in the "Preprints" as dealing with Area 7, there are in the volume a number of papers assigned to other areas which deal with the subjects under discussion here. Specifically: A Unified Index to Science (Area 2), by Eugene Garfield; International Cooperation in Physics Abstracting (Area 2), by B.M.Crowther; International Cooperative Abstracting on Building: An Appraisal (Area 2), by A.B.Agard Evans; Cooperation and Coordination in Abstracting and Documentation (Area 2), by Otto Frank; On the Functioning of the All-Union Institute for Scientific and Technical Information of the Academy of Sciences of the U.S.S.R. (Area 2), by A.I. Mikhailov; Recent Trends in Scientific Documentation in South Asia: Problems of Speed and Coverage (Area 3), by P.Sheel; Scientific Documentation in France (Area 3), by J.Wyart; and A Proposed Information Handling System for a Large Research Organization (Area 5), by W.K.Lowry and J.C. Albrecht.

He stated that the Panel, in its own discussions of the topic assigned to it, had identified the following questions as being those which lie closest to the surface:

What responsibilities for scientific documentation should lie with agencies supporting research?

Scientists (both as individuals and in groups)?

Scientific and other professional societies?

Abstracting and indexing services?

National documentation centers?

International documentation centers?

Industry?

How shall such activities be financed?

How can the needs for scientific documentation of the smaller or less developed countries best be met, and their contributions enlisted?

How can research looking to improvement in the techniques of scientific documentation best be promoted?

How is scientific documentation to be rationalized, simplified and unified so as to avoid scattering of effort?

How can training for scientific documentation work best be secured?

The Panel, he said, would attempt, with its limited membership, to touch on all these questions.

Dr. David C.Martin, Assistant Secretary of the Royal Society of London, speaking on the *responsibilities of scientific societies and journals,* commenced the Panel presentation by reminding the Conference that for three centuries scientific societies, through the publication of their journals, have been the principal builders of scientific information; they organize the supply of bricks and provide much of the mortar. He prophesied that for some years to come the paper published in the scientific journal will continue to be the principal medium through which scientific information is first dispensed. A large debt is owed, therefore, to the editors of scientific journals for their vital contribution. He suggested that this important service might be further strengthened by, among other things, the following:

a Societies should place greater emphasis on good writing—on the need for papers to be clear, precise, logical, and brief.

b Authors' synopses, subject to editing, could be helpful and should be further exploited.

c The encouragement of authors to indicate subjects for indexing and classification of their material as well as specific information for use in data processing might be useful.

d Teams of young scientists might be organized, perhaps by the universities, to extract information from the literature; there is no better way for getting acquainted with the shortcomings of the literature and for stimulating efforts for reform.

e Reviews should be written only by well qualified persons; more prestige should be given to this important information service. f Rapid announcement of titles after publication, followed by a speedy on-request service of individual offerings might be beneficial and should be tried.

g Societies should be financially independent and retain complete control of the publication of their original works. In return for the public benefit they give, the societies merit public support. Governments might provide funds to selected societies.

h The scientific unions can promote international coordination of the work of national associations.

i Continued emphasis should be given to the value of personal contacts.

j International agreements in such matters as nomenclature, transliteration schemes, etc. can provide national academies with a useful vehicle for communication.

Dr. Martin stressed the importance of patience and willingness to experiment. Too much should not be expected too soon, because cooperation on an international scale takes time. He anticipated the continued rise in the recognition of the importance of scientific information (this Conference will assist in this respect), even though the stage for the establishment of international scientific information centers might not yet have been reached.

Mr. Noble Clark, Associate Director of the Agricultural Experiment Station, University of Wisconsin, discussed the question of the *degree to which the scientists,* working individually or in cooperation with scientific groups, *should be responsible for contributing to the organization and control of scientific information* in forms generally available to the scientific community. He mentioned the suggestion made earlier in the Conference that the documentalist should give the scientist the information that he needs, rather than what he wants. Speaking as an administrator, Mr. Clark placed a high value on the expressed needs and desires of the individual scientist. Important new ideas in science, he explained, arise in the single mind of the scientist, whose convictions and preferences must not be ignored. In making a plea for finding new and better ways of taking care of the real and felt needs of the scientist, whether he works individually or in a research group, the speaker asked for continued efforts to find new and improved ways of access to the findings of scientific research in order to be of more help to the individual scientist. These methods need not always be elaborate; just as in transportation we have the great ocean liner and the one-man automobile, so in information work we undoubtedly need the computer-retrieval mechanisms for great team research projects, but we also need simpler devices to serve the needs of individual scientists, such as the following:

a A master card file of abstracts kept up-to-date in the institutional library.
b Duplicate cards for abstracts of interest distributed to individual scientists within the institution.
c Index words supplied with each abstract (perhaps by the wavy lines used to specify bold-face type).
d Cards adapted to machine sorting when files reach a size warranting such a procedure.

Sir Herbert Howard, Secretary, Commonwealth Agricultural Bureau, speaking on the responsibilities of the indexing and abstracting services, spoke of the differences in the needs of the various scientific disciplines and the degree to which these needs can be expected to be met by further cooperative effort. In any question involving cooperation, it is the scientist-users of the abstracts who must be the final judges of the efficiency of the operation. Discussion of means and methods should never be allowed to confuse the real objective of satisfying the needs of the user. The cooperative efforts that have been undertaken in physics and chemistry owe their success to author abstracts. These may be satisfactory in physics and chemistry, but biologists may need some very different kind of aid, particularly those who work in isolation without easy access to libraries or to literature. Touching on the matter of the high cost of specialized abstracts, Sir Herbert spoke of the sometimes unequal sharing of the cost in cooperative ventures, when one service makes use of the expensive abstracts produced by another without adequate compensation, sometimes even using them for competition. In summing up his principal points, Sir Herbert emphasized that:

a. Abstracts which are subsidized by governments do cost money; this money comes from the taxpayers,
b. Schemes of cooperation which satisfy the physicist or chemist are not necessarily suitable for the biologist,
c. In formulating a cooperative policy, those agencies that have produced abstracts for a considerable period should be consulted,
d. Inasmuch as specialized abstracts are expensive, any form of cooperation should provide benefits both ways,
e. The success of any cooperative scheme must be judged by those for whom the information is provided, rather than by those who provide it.

Dr. W.O.Baker, Vice President for Research, Bell Telephone Laboratories, discussed the *responsibilities of industry for contributing to the organization and control of scientific information* in forms generally available to the entire scientific community. Suggesting that industry has the important responsibility of communicating back to the scientist information on industry's use of and experience with science, Dr. Baker advanced the idea that industry should take responsibility for recruiting personnel capable of preparing in scientific terms reports on the industrial applications of science. Industry has responsibilities, he emphasized, not only as a user of science, but also as a critic on the application of the science. It is industry's job to extract what is valuable from the application of the science and it can serve a useful role by producing authoritative monographs, cogent surveys of its experience as a user of science, referring to original sources and written in scientific terms.

Dr. Burton W.Adkinson, Head, Office of Scientific Information, National

Science Foundation, speaking on the question of what responsibilities should rest with national governments, based his remarks on the following assumptions:

a Scientific information is a national resource.
b National governments are users of scientific information.
c Many national governments are responsible for the production of much of the scientific and technical information that is distributed today.

Recognizing the existence, even within the Conference, of completely opposite points of view as to the value of governmental control (e.g., as expressed in the papers by Dr. Milton O.Lee and Miss Hazel Mews), Dr. Adkinson enumerated several points that must be taken into account if national governments are to assume an important responsibility in this area:

a National governments must recognize their responsibility for the establishment and maintenance of adequate documentation services.
b National governments must provide a proper climate for the maintenance of adequate documentation services and accept responsibility for supporting and stimulating such services (this would involve whole-hearted cooperation between those who produce the information and those who organize and disseminate it).
c National governments will have to accept responsibility for fostering international cooperation in this field, since no nation is self-sufficient.

A variety of factors influence the role of the different governments:

a The stage of research and development programs both within and without the governmental structure.
b The stage and development of documentation activities within a particular country.
c The degree to which the country is a heavy contributor in scientific fields.
d The cultural and economic patterns of the country.
e The degree to which the country has developed bibliographical tools of international importance.
f The linguistic competence of the specialists within the country who use these major bibliographical tools in science and technology.

In summary, Dr. Adkinson emphasized the fact that national governments have been in the documentation business since the beginning of history. They must continue to be much concerned with this area. National governments have administrative and intellectual responsibilities as well as financial responsibilities here. How far each country can go in supporting documentation serv ices will be determined on the basis of such factors as those enumerated above.

Dr. C.E.Sunderlin, of European Research Associates, S.A., Brussels, in considering the *responsibilities that rest with both governmental and non-governmental groups at the international level,* referred to the recommendations for the establishment of international scientific documentation centers contained in the papers of Paul Boquet and Waldo Chamberlin. He stated that among the considerations bearing upon the establishment of such centers are: *(a)* measurement of the effectiveness of existing services; *(b)* determination of the existence of needs beyond what present services provide; and *(c)* considerations of the kind of organization required to provide such additional services. In his following remarks he limited himself to two questions, the first being, “Is there a need for scientific information services of a kind which would be persuasive of the establishment of an international center?” He believes that Dr. Boquet and Dr. Chamberlin have not made the case for such a need. Just to say that science is international is not to demonstrate a need for an international center —art and sex are both international. He would greatly question whether the information services now being provided at the international level by various services could be rendered more effective if they were provided by a single international center. To the question, “How can existing needs best be met?” he would say that scientific information services must ultimately rest on great scientific collections, and that this argues the establishment of a series of national and regional science libraries functioning as national and regional science centers. The proposed National Lending Library for Science and Technology in the United Kingdom, and the various regional science documentation centers established under the sponsorship of Unesco are harbingers of such a system. Such centers would be coupled with continued growth and experimentation in international activities for specific areas of science, and questions of sponsorship, financing, policy determination, organization and control of services would be solved on a case-by-case basis.

Dr. Juan Carlos Secondi, Director of Libraries and Public Information, Provincial Health Service of the Chaco, Buenos Aires, outlined some of the problems arising from the high rate of illiteracy in the world today. Explaining some of the attempts being made in Latin America today, including the contributions made by Unesco, to lift the barriers to learning, Dr. Secondi saw technical assistance as perhaps the best tool to progress. The less-developed countries need help from the more-developed, in scientific documentation as in other matters, until their level of development is such that they can go ahead alone.

Dr. Jesse H.Shera, Dean, School of Library Science, Western Reserve University, next spoke on the matter of training for work related to documenta tion. He stressed the importance of the maturity that comes to a profession from experience and the fact that the field of scientific documentation is still very young. While he doubted that there was any one ideal school curriculum, Dr. Shera viewed education as the bridge to the future. Although one cannot expect a research program in education to find out for certain what kind of training is best, the maturity that will come only from experience will help to indicate the most suitable kinds of training for this field.

Dr. Raymund L.Zwemer, Assistant Science Adviser, U.S. Department of State, reviewed the principal steps to be followed in furthering research in the communication of scientific information.

a Find the man with the idea that needs to be developed and give him backing,

b Make certain that those working on related subjects have an opportunity to get together for discussion, mutual aid, and constructive criticism,

c Provide adequate publication outlets for exchange of information,

d Develop in a practical area but on a small scale the most promising of these ideas.

Dr. Zwemer stressed the importance of controlled studies of various methods of documentation, so that they may be compared in similar terms (this is seldom now the case); he also stressed the importance of personal communication as a medium for the transmission of scientific information, the need for better studies on the scientists' own use of informational materials, and the need for more investigation on the integration of research. He saw considerable merit to the suggestion that the scientist might put in the headings and index words under which information in a published work might be recovered. Systems compatibility on the international scale will become increasingly important. Governments must look carefully into the language problem, for example, and into the financial barriers to the free flow of scientific information. In the international field, it would be helpful to have a description of the techniques followed in the regional centers which are just beginning to grow into prominence. A better assessment is needed on the magnitude of what can and should be done in the light of what we can afford to do.

Discussion from the floor brought out a number of points.

GENERAL

Dr. D.J.Urquhart, Department of Scientific and Industrial Research, London, commented that the papers in Area 7 appeared to be deficient in basic data, and rather concerned instead to put forward personal views. He himself had been inclined to prepare a paper in this area, on an inquiry of some consequence, but had felt that his conclusions lacked the basic character which should have been a prerequisite for papers for this Conference. He had observed that English abstracting journals deal with scientific literature in English better than do the Soviet abstracting journals, and the Soviet abstracting journals better with Soviet literature than with English; this might have been expected. He then found, however, that Soviet abstracting services deal better with the German literature than do the English services, and this seems to be due to deficiencies in collections of the German literature in England. As a consequence, the National Lending Library for Science and Technology is being planned in the United Kingdom. But this does not result from a conclusion that a centralized system is better than a decentralized, but rather from observation of operational, as opposed to fundamental, aspects of organization.

Mrs. Helen L.Brownson, Office of Scientific Information, National Science Foundation, pointed out that the comments of Mr. Clark, Dr. Sunderlin, Dr. Boutry, and Dr. Urquhart all add up to the fact that we just simply do not know enough, and that what we need is real work, more studies in a number of the areas discussed during the Conference.

Professor Eric de Grolier, Director, Centre Français d'Échanges et de Documentation Techniques, Milan, warmly seconded Mrs. Brownson's suggestion, adding that there must be a series of specialized conferences or symposia at which real research projects in the field of information communication could be devised. For “even Hell has its laws,” as the French say, and the hell of scientific documentation may have them too.

Professor J.D.Bernal, Birkbeck College, University of London, stated that finance should not become an obstacle to improvements in the dissemination of information. The centralized service could begin as a free service, with the government supporting it. Or it might be sold to industry, with the resulting revenue used to finance research in the problems of dissemination. Perhaps a finance committee is needed to look into the problem of the degree to which high cost, including the increasing cost of books and journals, is a barrier to communication.

COORDINATION OF EFFORT IN INDEXING AND ABSTRACTING

Dr. G.A.Boutry, Secretary, Abstracting Board, International Council of Scientific Unions, commented that no attempt has been made to assess the magnitude of the problem with which we are dealing. He notes, however, that out of the mass of abstracted literature, less than one fiftieth is devoted to mathematics. Between one fifth and one sixth is devoted to physics, and less than one third is concerned with chemistry. All the rest, about 50 per cent, is devoted to biology and its allied disciplines. These rough statistics would seem to explain why there are so many problems in the documentation of the biological disciplines, and so few, comparatively, in mathematics. It explains, too, why there has been little difficulty in the attempt of the ICSU Abstracting Board to organize the field of physics. As for principles, the Board has none; it forces no one to do anything. But when it started, the only agreement the various parties could reach was that authors' abstracts, even though they were not thought to be any good, yet cost nothing and so were worth trying. But it is not to be expected that they will meet everybody's needs. So, the Board is willing to compromise in an effort to secure action which will be some good to some, without expecting that it will find the perfect solution for everyone.

Mr. Cyril Cleverdon, College of Aeronautics, Cranfield, Buckinghamshire, said that it seems sensible to ask the writers of a document to indicate the index terms under which the information therein can be found.

AN INTERNATIONAL SCIENTIFIC DOCUMENTATION CENTER

Dr. Paul Boquet, Institut Pasteur, Paris, remarked that man has been concerned with the retrieval and dissemination of information for thousands of years. Even though much progress has been made, the problems remain complex, partly because of the difficulty in standardizing methods and systems.

Dr. Waldo Chamberlin, New York University, stated that after five days in the Conference he felt that his paper was too conservative. Since so many people are apparently doing the same thing in the field of storage and retrieval of scientific information, a modest goal might be the establishment of an informational center aimed at improving our present methods of storage and retrieval. Present costs of decentralized services appear to be higher than would be required by an international center. But irrespective of the cost, can we afford the present wasteful effort? The Soviet Government has decided that it cannot afford it—perhaps other countries cannot afford it either. This Conference can undoubtedly make a concerted attack on the problem if it chooses to do so. In politics we get what we want.

Dr. Masao Kotani, Chairman, National Committee for Documentation, Science Council of Japan, and Dr. Wallace W.Atwood, Jr., Director, Office of International Relations, National Academy of Sciences (U.S.A.), forwarded the idea that international centers are perhaps the most sensible answer to many present needs. The experience of the three world data centers collecting information during the International Geophysical Year offers a precedent which should be watched carefully to see if it may offer a solution to the problems which have been considered at this Conference.

THE ROLE OF THE SCIENTIFIC INFORMATION OFFICER AND TRAINING FOR DOCUMENTATION WORK

Mr. J.E.L.Farradane, Tate and Lyle Research Laboratories, Keston, Kent, and Dr. Mortimer Taube, Documentation, Inc., Washington, D.C., brought out that it does not give an accurate picture to say that the documentalist is not competent to know what the scientist wants and needs. Many documentalists have been research scientists and thus have a high appreciation of the scientist's needs. The entire field of scientific information services should be recognized as a newly organized professional activity that will help both the scientist and the public and that merits recognition by both.

Mr. Clark interjected a request that documentalists not become too defensive or take themselves too seriously.

With respect to training for documentation work, Dr. S.R.Ranganathan, Professor of Library Science Emeritus, University of Madras, stated that the similarity between the work of the librarian and the information officer would seem to argue for similarity in their training. The curricula of library schools in India are now being used effectively for the training of people in both professions. He concluded by asking the Conference not to confuse issues by drawing unnecessary and non-existing distinctions between librarians and information officers.

Mr. Farradane agreed heartily with Dr. Ranganathan, and stated that the paper prepared by himself and Dr. A.B.Agard Evans for Area 7 is out of date in this respect.

SIMPLIFICATION AND UNIFICATION OF SCIENTIFIC INFORMATION SERVICES

In bringing the session to a close, the Chairman mentioned that perhaps insufficient attention had been given both in the papers and in the discussion to the question of what is needed to provide greater unity in documentation research in order to overcome the present scattering of effort. This desire to find a simplified method is in some respects the natural result of human laziness, and while there are many precedents there are large economic problems to be faced. We do not know with any precision what the boundaries of the scientific information problem are in economic terms. In April 1953 *The New York Times* reported that the cost of literature searching in the United States had been estimated at some $300,000,000 a year. Dr. Chamberlin has estimated that about $200,000,000 annually is going into the compilation of the publications that are searched by these literature searchers. Unquestionably duplication is a basic problem here, as is also the enormous growth of research institutions. Further economic studies appear to be an urgent need, and some funds ought to go to research toward the discovery of satisfactory methods of simplification.

VERNER W.CLAPP, *Rapporteur and Discussion Panel Chairman*

FRANK B.ROGERS, *Area Program Chairman*

CLOSING SESSION

Summary of Area Discussions

Members of Discussion Panel

Chairman: MILTON O.LEE, Federation of American Societies for Experimental Biology, Washington, D.C.

HELEN L.BROWNSON, Office of Scientific Information, National Science Foundation, Washington, D.C.

VERNER W.CLAPP, Council on Library Resources, Inc., Washington, D.C.

ERIC DE GROLIER, Centre Français d'Échanges et de Documentation Techniques, Milan, Italy

SIR ALFRED EGERTON, Salters Institute of Chemistry, London, England

DWIGHT E.GRAY, Office of Scientific Information, National Science Foundation, Washington, D.C.

ALEXANDER KING, European Productivity Agency, Paris, France

GILBERT W.KING, I.B.M. Research Center, Yorktown Heights, N.Y.

JOHN W.TUKEY, Department of Mathematics, Princeton University, Princeton N.J.

Closing Session

At the closing session of the Conference the Chairmen of the Area Discussion Panels attempted to bring together the more important points that emerged in the Area discussions, whether or not there was consensus or disagreement on these.

Area 1

In general terms, the main needs of scientists are fairly clear. They need means of becoming aware of information of concern to them; they need access to particular journals and to copies of particular papers; they need means of access to the whole scientific literature both in their own and other fields; and they need up-to-date media such as monographs, reviews, and compendia.

In the discussion for Area 1 there appeared to be tacit agreement on several points.

1 The requirements of users of scientific information and the uses to which they put the information they require should determine the design of information systems and services.

2 Study of the informational requirements of scientists is difficult, for the communication of scientific information is extremely complex. Needs vary with the subject field, the type of research, the availability of information services and source materials, geographical situations, differences in human abilities, and the use that is to be made of the information. Furthermore, the real needs of scientists and their expressed wants may be quite different. It was pointed out that the studies made thus far have not been concerned with information requirements as such. They have sought data on what scientists do, what they think about the present situation, and what they say they want in the way of improved services. The development of information services based simply on an analysis of present user habits is likely to be inadequate.

3 The studies made thus far, although admittedly imperfect, have produced data and conclusions that have served as a basis for action in particular local situations. It was suggested that the studies already

reported be carefully reviewed, evaluated, and compared. These studies appear to indicate certain things:

a The various services now available to the research worker would serve to a considerable extent in meeting his requirements if he were more wise in their use, developed more facility in their use and made more use of them.

b Scientists place great reliance on face-to-face and other informal means of communication. Personal communication, however, and the conveyance of information through the literature are mutually dependent.

c The temporal span of the usefulness of the literature for current research is relatively short.

d Abstracts are less used than one might expect; information is obtained rather more from journals and the references they cite.

e Interdisciplinary communication is both important and difficult.

4 There is still need to know more about the real requirements of users, the various factors affecting the efficiency of scientific communication, and to develop more sophisticated and reliable techniques for studying them. Studies should be designed so as to produce conclusions that could be tried out experimentally. A number of suggestions concerning future work in this field were made:

a In future studies of the operations research type, a qualitative judgment as to the value of the information should be obtained along with quantitative data and objective measurements.

b Knowledge of human perception, motivation, capabilities and limitations is relevant to the whole communication problem, and the point of view of the physiological psychologist should be taken into consideration in planning future work.

c Socio-psychological inquiries, of the sort made by Bavelas, might be made of the nature of the conversational way of getting information. For example, one might set up a simply-patterned “game” involving information gathering in order to see how differently structured teams succeed.

d Inquiry into the scientific meeting as a means of transferring information should be made. One might ask: Are present meetings too large? Are ten-minute papers too short? Does one always learn more “in the hall” after the sessions than during the sessions—and, if so, why?

e Inasmuch as no research worker has access to all the information he ought to have, it is important to determine to what extent this matters.

f Since it is difficult to remember what one does and uses, perhaps it would be possible to devise a system of tallies to be placed on library-materials so that their use could be summarized.

g Other suggestions concerning future studies were made in their conference papers by Professor Bernal and Dr. Menzel.

It was pointed out in the Area 6 discussion that for a sound theoretical analysis of information retrieval systems a greater understanding of empirical processes in their use is needed; and that the person who uses a system machine should therefore be required to describe himself in a way that will indicate his field of interest and objective. Until a hierarchy or classification of users and their needs has been established, however, it is not at all obvious how we should ask users to describe themselves.

It was suggested that much of the problem of better and faster information services could be solved rather simply if the financial support for information services could be made to keep pace with the support of scientific research. Certainly it is clear that more support for our information services is needed and that this would solve some of our problems. The funds we can spend on information services, however, are not unlimited and the number of qualified persons who could be put to work on them is limited. Since there is a limit to our resources, it will always be important to try to make the allocation of these resources that will best meet the needs of scientists.

It has to be remembered that the body scientific is an organic whole adapting itself to circumstances and changing all the time. Conclusions drawn from studies may be to some extent ephemeral. Provisions for meeting information needs should therefore be flexible.

Area 2

In Area 2 there was considerable discussion on the advantages and disadvantages of highly centralized, comprehensive, national information services, versus the decentralized type where much of the activity ties in closely with professional societies.

The discussion on the question of the authorship of abstracts reiterated what has been said many times on author versus non-author preparation, indicative versus informative content, telegraphic versus complete-sentence style. If one can generalize from the limited analytical studies that are reported in the Conference papers, the information content of author and non-author abstracts differs very little; the author-prepared abstract is likely to be both a little more precise and a little more verbose. This seems to indicate that the intellectual job of preparing abstracts or annotated titles should be placed on authors, with some educational effort expended to achieve conciseness of expression. Some further studies in this aspect might end the discussion, hitherto based on opinion only, on author versus non-author abstracts and the alleged slanting of abstracts.

A point made in connection with comprehensiveness of coverage was that of what one is willing to pay for that little last approach to perfection. It is somewhat like the comprehensive bibliography problem where finding the last five per cent of the references requires 95 per cent of the cost. At some point one has to ask, is it worth it? Even if we postulate that the best possible abstract for a given scientist is one written by another expert in that scientist's own narrow field of interest, what would one have to pay in terms of money and time to have this? Is it worth it, or should we compromise in certain other ways in order to get a more generally valuable answer?

There was a considerable amount of favorable opinion for a sequence that would consist of an author draft written according to a generally circulated set of ground rules, followed by the editor of the primary journal taking the same kind of responsibility for the abstract that he normally does for the paper. He should feel perfectly willing to send the manuscript back and say the paper is all right but the abstract has to be done over again for such and such reasons. The abstracting service editor, of course, would have the final veto or modification power over the result.

Another point noted had to do with cooperative efforts beyond the one of exchange of proof of abstracts among various services. As it stands now, major abstracting services in the subject fields use different classification systems, different indexing systems, different abbreviation lists for journals, different translation assistance. Certainly this is an area for cooperative effort, and such is already in progress—in the field of physics and chemistry under the ICSU Abstracting Board.

The discussion in both Areas 1 and 2 strongly emphasized the need for better education of users and training of their habits in information search and retrieval. This training should begin at least in the graduate school. The design and development of information services should be determined in part by the present habits of users and in part by habits that might reasonably be expected to be conditioned by training.

It was pointed out that our present type of information services are at once highly developed for the fields of applied science or technology (medicine, agriculture, metallurgy, engineering) and at the same time most inadequate for satisfying urgent needs to know in those fields. The inadequacies of indexes are particularly serious.

The opinion was expressed that the present organization of uncoordinated abstracting and bibliographic services is on the verge of collapse, and that it may have to be replaced soon by another type of organization, based on specialized centers, each retaining a considerable degree of independence but combined and harmonized in an over-all system. Such a system would almost inevitably be a highly mechanized one and could be used to full advantage only if the principles of division of labor, coordination and large use output are applied.

Area 3

In Area 3 the discussion of the effectiveness of monographs, compendia, specialized centers, and of new techniques and types of services led to no detailed consensus; indeed none could be expected. Some generalizations did emerge, however.

The present rapid growth in the mass of scientific information, in both its fundamental aspects and its utilization, in the scope which it covers and in its complexity, parallels exactly present political and economic growth. No longer can any countries except the very largest ones be really self-sufficient in politics or economics or science.

It is just the same with scientific information. We can no longer be self-sufficient either as individual scientists, as institutes, or as countries. With the increase in the quantity of publication and the ephemeralness of its importance, there is occasioned a tremendously increased need for selection of what is significant, and speed in getting it to the individual scientist user. With the increase in the amount of specialization and subspecialization in science, there is need not only for the subjects discussed in Areas 1 and 2, namely, indication of the existence of information and availability of the original, there is much more need also for complementary scientific activities and background scientific publications such as reviews, compendia, monographs, and so forth.

No one asks for fewer of these, although there was some doubt in some of the papers as to whether large systems like Beilstein and Gmelin could in fact continue, because they are very expensive. The answer is probably they will have to continue.

There is also agreed a need for annual reports on the advancement of this or that subject, for monographs giving comprehensive treatment of individual documents from time to time as their importance becomes clear, and a very great need for critical evaluation. We have heard much in the last year or two about the importance of the scientific literature from the Soviet Union. A great deal of translation is being done. But one gathers the impression from talking to working scientists that the real user's demand for this material is not yet fully developed or is likely to be for two reasons. Much of the material has not been available because of language difficulties until now. Also, people are uncertain as to the value and the perspective to be given to the literature of a country with which they are not completely familiar. Therefore, the critical type of review or monograph which brings forward and presents a number of the important points of growth in a science has become especially valuable in recent years, and will become more valuable as it becomes more and more difficult to read anything except a very small portion of the papers on subjects in which one is interested.

In addition to this type of non-serial publication, research and laboratory reports have an immediate though transient importance because of their speed factor. Although these complicate the system and although they may not be abstracted completely because eventually the material they contain turns up in papers in journals, nevertheless there is a need for them.

Compendia, critical tables, and quantitative data of all kinds, which represent the concentration of scientific data from the past, may be the most important type of scientific information of all. But even the accumulated quantitative data are growing in amount so quickly that summaries of the more important and more commonly used data are also required. There is clearly a great deal to be done in this field which can be classified generally as non-serial publication.

The situation with respect to special information centers and services is very different for fundamental science from what it is for applied science. In fundamental science, a national or an international specialized institute or information service has very often as its function the publication of original material as well as the production of informative abstracts, reviews, and bibliographic materials of all kinds.

On the whole, there is far more need for the specialized information center in applied science. In the Conference the discussions very properly concentrated on the scientists' need for scientific information, and have said rather little specifically about the need for information by the practitioners of applied science, although the economic compulsion here is the reason why so much science is being done. In applied science for industry, agriculture, medicine, and engineering, information requirements are of two types. There is first the searching of the fundamental scientific literature to find what is appropriate for the particular field of application, and, secondly, the additional task of finding out what is new and significant in application and development for that field.

So the task here is different, and decentralized activities, although connected in a centralized system, are inevitable. For example, in a rapidly developing industrial field the progressive firm needs a great amount of both basic and applied scientific and technical information as rapidly as it becomes available. The technical information may be fairly easy to gather from a limited number of journals. For the fundamental scientific advances on which all of the technology rests, there is a minute proportion of pertinent material in each of a great number of journals. Consequently the specialist information center with an industrial objective has to vary and multiply the tools required in information work. It has probably a great need to provide for industry abstracts of a completely different type from what is the basic service need for indicative abstracts. It may in fact abstract from a certain paper only one paragraph which is relevant to the particular industry, but the informative aspect of that abstract may be of tremendous importance.

Consequently many of these bodies bring out and publish after much searching, informative abstracts of a particular type. They also encourage among their firms and among scientists in their industry verbal communication through such media as symposia, conferences, and of late there has been a great growth in many parts of the world of field liaison services of various kinds.

Area 4

In the concluding survey of Area 4, the discussions were concerned with the scope and difficulties of comparison of systems, the need to determine primary objectives for any retrieval system both for purposes of design and of evaluation, and with the question of determining the specific criteria to be used in testing the merits of a given system.

Comparison and evaluation of systems are still very hazardous, because of the many variables that enter into the design of any particular system and the fact that the design of any system is decided to some extent on specific objectives and somewhat unique combinations of objectives. It is doubtful that the comparative tests that have been made or that are planned can decide the relative merits of systems for general application. They can only help to decide on the merits of certain particular systems for certain particular applications. Hence there is real danger in drawing general conclusions from experience that is as yet too limited.

The observation was reiterated that our new machines, systems, and methods have only supplemented and not replaced the old. As various systems in scientific communication have evolved, there is a continuing coexistence of the previous ones, from oral communication, correspondence between working scientists, publication in primary journals, secondary media of information, bibliographical aids, and the present multitude of large and small documenta tion centers for at least the functions of accessioning, processing, and redistributing material. Is it likely that this pattern will be reversed and that various mechanisms developed in the past will be supplanted completely?

Areas 5 and 6

In summarizing the discussions in Areas 5 and 6, it was reiterated that despite claims of proponents, none of the systems in use or suggested is entirely attractive and safe from the point of view of the scientist-users themselves. It is not the nature of machines that is going to determine the systems, but the nature of the scientists who use them. Even in systems of indexing classification where there is always a certain amount of interplay between the seeker and the librarian, documentalist or machine operator, the basic frame of reference must be the seeker. This second stage of interplay is still highly unorganized. It is hoped that impetus will be given by the Conference for undertaking serious efforts in operations research study of types, forms, vagueness, degree of sophistication and specificity of questions by users.

Another type of problem upon which some people will be interested in doing research in the next few years was touched on briefly in the discussions. It is not too difficult to make a retrieval system work for a small or circumscribed body of knowledge. The difficulty arises with the overlaps into different fields. New problems arise and it is not yet clear what these problems are. To bring enormous systems into existence, with the rigidity that their existence implies, before these problems are thoroughly recognized and explored would be a mistake.

Many fields and attitudes were represented at the Conference, whose greatest value may have been through increasing the appreciation of these complexities and of the relevance and attitudes and actions of other groups. Thus more librarians have come to understand that the mathematical formalization of procedures that librarians have long carried out competently and informally is a necessary prerequisite to the automation of even some of the easier aspects of a librarian's work. Information retrieval is, like all other broad fields, in the process of mathematization.

THEORY

It seems obvious that theories must grow by stages and that a complete general theory of information may never come. It is also obvious that readymade mathematical techniques are unlikely to take us very far. Mathematical theories need to be developed within a wider framework of functions, that is, functions in the sense of the way in which things work, not in a mathema tician's sense. Here one wide enough to cover systems both present and future, both manual and automatic, is needed.

For the next few years such theories may be expected to have limited generality. But, for all of this, mathematical theories developed with the aid of deep insight and experimental verification is the long-run hope of information retrieval. In particular, the applications of information theory to information retrieval will be specialized, possibly including, for example, such things as the optimum arrangement of documents in a file not subject to direct human access or the optimum coding of addresses for such a file.

MODELS

Further work on over-simplified mathematical models carefully evaluated will be valuable and should be actively continued. The results thus obtained should be regarded as starting points, not stopping points, and should be compared with experience and experiment to show which of the considerations left out is the most important.

EXPERIMENTATION

A wide variety of types of experimentation is badly needed. Arrangements of physical objects, trial libraries, simulations on general purpose computers, even writing experimental computer programs, all are experiments. Adequate attention must be given to problems of scale, particularly to the use of extreme ingenuity in modifying small scale situations so that they will exhibit behavior ordinarily shown on much larger scales. In particular, objective and empirical studies of present-day patterns of both queries and retrieval are needed. It is seriously to be doubted if we know how retrieval is actually being carried out with manual systems.

INFORMATION RETRIEVAL SYSTEMS AS WHOLES

There is a growing realization that a large mechanized system will include functions which today's manual system assigns to the library staff. However, information retrieval systems need not be single stage even after mechanization. Breaks between stages may correspond either to human intervention and decision or to alterations in a programmed strategy. Indeed, the problems of large, medium and small systems differ enough to generate quantitative differences in the character of optimum solutions likely.

INFORMATION RETRIEVAL SYSTEMS IN RELATION TO PEOPLE

It may well prove desirable to ask persons connected with in-put, that is, authors, referees and so forth, to aid information retrieval by simple expressions of judgment. We must attach more importance to the continuation of both freedom of expression by users in querying, even though many queries may be maladroit, especially since old habits of query formulation will at best die slowly, and education by the information retrieval system of those who ask queries. For these and other reasons, two-way communication between querist and information retrieval system, human or machine, will continue to have high priority. We sometimes forget that we do have two-way communication between the querist and the librarian. Indeed, as such strategies are automatized, users are likely to have to teach the machine how to improve its search strategy.

INTERNAL SEARCH, LANGUAGES, CLASSIFICATION, AND RELATED TOPICS

The language in which queries are discussed with the querist need not be the same as the internal language in which a query expresses the desire for search. Internal search languages of three extreme types received the most discussion.

1 Hierarchial ordered languages, that is, classification in the restricted sense.
2 Languages consisting of descriptions, simple, inflected, combined with relations, and so on. About these, two remarks are important. Such relations as interlocking and interfixes are likely to be but the beginning of relational descriptive syntax. The presence of either inflexional or relational descriptive syntax requires mathematical structures more complex than those of lattice theory.
3 Associative languages in which expressions of the relevancy of pairs of documents provided the basis for search and queries are expressed in the form “like this and this and this, but not like that and that.” Again, two remarks are needed. Classification is implicit in such a system, and thus much more easily automatically up-dated. Second, the documents are likely to be parts of present-day documents. These simple extreme types were thought of only as exhaustive. The need for including the querist's background, for example, as “an organic chemist interested in dyestuffs,” was stressed, as was the possibility of simultaneous use of two or more sublanguages as an elementary, not very realistic example, as if the UDC and the Colon classification were used simultaneously. In mechanization of classification, internal search language must be specified as a tool, not as an object of art, although semantic considerations need not be absent.

SEARCH STRATEGY, MACHINES, PROGRAMMING

Details of search strategies are ever less clear than details of internal search languages, but the view was strongly expressed that such strategies must be closely tied to the most modern techniques for programming digital computers, techniques which go far beyond computation. Today programs which involve randomness of start and guidance by results, so-called heuristic programs, beat good players at checkers and have given neater proof of one theorem of propositional calculus than did Whitehead and Russell. Thus self-checking exploration allows machines to handle a new group of actions once considered “thought.”

Special formal languages used in programming are now so flexible and powerful that they may be much better adapted to the discussion of search strategies than either natural languages or the formal languages of pre-programming mathematics. Techniques for obtaining programs which may be safely modified by intervention without detailed study (and hence may be safely self-modified) appear to be in sight. Both heuristic and safely self-modifiable aspects are likely to be essential if search strategies are to be both teachable by human intervention and adequately able to profit from experience. These are both abilities which will continue to be of great importance in information retrieval systems.

RELEVANCE AND ASSOCIATION

The problems of using a specific formula to express mutual relevancies (query to document) were discussed and many ideas were hinted at. There is a clear need for both cooperative and collaborative work to provide several reasonable alternatives and for some real trials. Similar needs exist in other areas of search strategy where many problems call for combination of background, for careful discussion and for experimentation.

POST-SEARCH PROBLEMS

Once search procedures have determined the call-signs of documents, abstracts, and so forth, there remain significant engineering problems of how best to arrange for the fetching of these documents or abstracts.

Some things that have been mentioned are but hopes. Others have already been done, but without regard to cost. The ideas are stimulating and the net value of their discussion and consideration is great so long as we remember that many of them will not be embodied in procurable devices for a substantial time. Meanwhile, those with immediate problems must do the best they can now. They dare not wait for an overall general theory whose approval is unpredictable nor for theories of intermediate generality, nor for safely modifiable heuristic programming, which will come sooner than most of us think. New and diverse special and interim solutions can and should be attempted. They will contribute both to the solution of the specific problems and to our common store of knowledge.

Area 7

In summarizing the discussion in Area 7 there was a consensus that this area shared responsibility in the entire subject of the Conference, although the responsibility is not of a specifically scientific sort. It is rather at the administrative level but it invokes there the arts and politics. It involves the support of documentation services, the support of research in documentation and the whole question of the practicability and the method for working toward unification and simplification.

There is as yet no measure of the problem which we are facing, no complete picture of the present economics of the situation, and not even any good descriptions of methods employed to work out present arrangements in economic and administrative terms. We do not even have a systematic catalog and description of the tools that are now available. Yet the real reason for the Conference, beyond the mere fact that we are faced by mountains of documentation, is that we do have available some very powerful tools, specifically that group of techniques, arts and sciences which are lumped together under the name automation. There is the existence of the computers and the other data processing machines, of material moving devices, of mechanisms for reducing to minute facsimile and to other terms and in other ways the facts and information in which we are interested, and for transmitting information at high rates of speed. The powerful tools of mathematics, of statistical analysis, and the techniques of the behaviorial sciences, including psychology and that group of techniques brought together frequently under a name such as operations research, the tools of linguistics, all these are around us in active development but are to a very limited extent brought to bear on this particular problem.

There was tacit agreement that full use must continue to be made of all the facilities, of all the organizations working in this area, and of all the potentialities which are available for improving work. Specifically, as for the responsibilities of scientists individually or in groups as scientific societies, of the indexing and abstracting services, of national governments and of international groups, both governmental and otherwise, there seemed to be no inclination on anybody's part to say that specific responsibilities of one kind or another always adhere to one or another of these bodies, or that any of them can now be relieved of any responsibility which they now have undertaken.

It also seemed to be accepted that attempts at global solutions should be regarded with some suspicion and question. Progress is to be made rather by small experiments or, as one speaker put it, by seeking answers in restricted areas, permitting us to walk before we run.

International cooperation must certainly be encouraged, but in its early stages between national groups rather than on a global international centralized scale. Many types of continued research and inquiry are needed on the size of the problem, on the difference of the size of the problems as affecting the various sciences, the various different disciplines, on the effects of differences in size on the kinds of documentation services and the organization of these services, on the economics of the problem, and on the real reasons for the failure of certain efforts.

In the field of education for scientific information work, it seemed to be agreed that interesting developments were in existence worldwide and that there is no need at the present time for any freezing of the situation. Rather, it is preferable to let the present experimentation proceed and get where it can. In the matter of unification and simplification, too little attention was given. It was noted here again that the tremendous cost and expenditure of effort which the present situation represents poses the question: can we afford to waste this energy in the present scattering, duplication and laggardness of effort?

It might be agreed that this question should be answered in the negative, but as to what action should result therefrom, there was no consensus. There were sour looks at all the proposals for centralization, but this entire picture might be changed by the change of one detail in either technological or methodological development.

A basic problem is that of organization, including integration of services both nationally and internationally, and of negotiation between organizations and countries as to what shall be done and in what manner to meet the needs of different kinds of scientists and institutions much more carefully and precisely than hitherto.

This organization, including a division of tasks, provision of meeting specialized needs, and the establishment of a certain degree of standardization may be all that is required, except financial support. Certain clearing-house activities and communication services are the maximum that might be necessary for world centralization or even national centralization.

As to standardization, component rather than rigid standardization is the desirable goal. We need the components in different organizations and in different countries to be of such a type that they at least fit together. This, rather than the external form is what matters, in fact the variety which is the richness of science must be preserved. The problem is essentially one of sociological, organizational, and administrative type. Although it is complex, it is no more difficult than other problems that man as a social animal has already solved.

and Area Chairmen addressed themselves particularly. Many specific suggestions, partial solutions and new points of view in addition to new factual information were presented during the Conference and in the working papers. It was urged that these not be neglected, but that they be worked over, digested and studied in and through a continuing activity of the Conference. Rather than waiting a number of years before holding another large general conference, it was urged that beginning in the near future there be organized a series of smaller symposium-conferences on restricted and specialized areas and particular topics, both those of the subject areas of this Conference and new ones. These should be organized on a more truly international basis than was possible for the present occasion. Without a deliberate follow-up much of the value of the Conference would be lost.

MILTON O.LEE, *Rapporteur and Discussion Panel Chairman*
CHARLES I.CAMPBELL, *Program Committee Chairman*

FINANCIAL SUPPORT

Financial support from the following is gratefully acknowledged.

UNITED STATES GOVERNMENT

National Science Foundation

Air Force

Office of Scientific Research (ARDC)

The National Institutes of Health

United States Public Health Service

Department of the Army

United States Atomic Energy Commission

United States Department of Commerce

National Bureau of Standards

Office of Technical Services

Patent Office

Weather Bureau

Department of the Navy

Office of Naval Research

Department of Agriculture

Agricultural Research Service

PRIVATE ORGANIZATIONS

Abbott Laboratories Library

Allied Chemical & Dye Corporation

Allis-Chalmers Manufacturing Company

American Can Company

American Cyanamid Company

Archer-Daniels-Midland Company

Bell Telephone Laboratories, Inc.

Bethlehem Steel Company, Inc.

Boeing Airplane Company

California Research Corporation of the Standard Oil Company of California

Cities Service Research and Development Company, Inc.

Council on Library Resources, Inc.

The Dow Chemical Company

E.I. du Pont de Nemours & Company, Inc.

Eastman Kodak Company

Esso Research and Engineering Company

Food Machinery and Chemical Corporation

Ford Motor Company Fund

General Dynamics Corporation

General Foods Corporation

Hercules Powder Company, Inc.

International Business Machines Corporation

Kennecott Copper Corporation

Lockheed Aircraft Corporation

Mercke Sharp & Dohme Research Laboratories Division of Merck & Co., Inc.

North American Aviation, Inc.

Chas. Pfizer & Company, Inc.

Pittsburgh Consolidation Coal Company

The Procter & Gamble Company

The Rockefeller Foundation

Shell Development Company

Smith, Kline & French Laboratories

Socony Mobil Oil Company, Inc.

The Standard Oil Company of Ohio

Sun Oil Company

The Texas Company

United Aircraft Corporation

United States Steel Foundation, Inc.

EXHIBITORS

Bendix Aviation Corporation
Benson-Lehner Corporation
Documentation Inc.
Eastman Kodak Company
Gallant, Inc.
Eugene Garfield Associates
George Washington University
Information for Industry, Inc.
International Business Machines Corporation
International Federation for Documentation
Jonker Business Machines, Inc.
Kalvar Corporation
The Microcard Corporation
National Science Foundation
Rabinow Engineering Co., Inc.
Recordak Corporation
The Royal Society of London
The Science Press
The Service Bureau Corporation
Smith, Kline & French Laboratories
Société Ceramique et Méchanique (Department Filmorex)
United States Atomic Energy Commission
United States Book Exchange, Inc.
University Microfilms, Inc.

ROSTER OF REGISTRANTS

A

Abelson, Philip H.
Geophysical Lab.
Carnegie Institution of Washington
2801 Upton St., N.W.
Washington 8, D.C.

Ackoff, Russell L.
Operations Research Group
Case Institute of Technology
Cleveland 6, Ohio

Adams, Scott
4621 High St.
Chevy Chase 15, Md.

Adkinson, Burton W.
Office of Scientific Information
National Science Foundation
1951 Constitution Ave., N.W.
Washington 25, D.C.

Albrecht, J.C.
Murray Hill Lab.
Bell Telephone Labs., Inc.
Murray Hill, N.J.

Alexander, Samuel N.
Data Processing Systems Div.
National Bureau of Standards
Washington 25, D.C.

Allen, Layman E.
Yale University Law School
New Haven, Conn.

Allison, Nancy Jane
Technical Library
Callery Chemical Co.
Callery, Pa.

Altman, Berthold
Ordnance Corps
Diamond Ordnance Fuse Labs.
Washington 25, D.C.

Amarel, Saul
Radio Corp. of America Labs.
Princeton, N.J.

Ammannito, Theresa
Allegheny Ludlum Steel Corp.
River Rd.
Brackenridge, Pa.

Anderson, Edwin J.
National Science Foundation
1951 Constitution Ave., N.W.
Washington 25, D.C.

Anderson, Hattie T.
Grace Research and Development Div.
W.R.Grace & Co.
Washington Research Center
Clarksville, Md.

Anderson, J.W., Jr.
Industrial Div.
The Magnavox Co.
Fort Wayne 4, Ind.

Andrews, Don D.
Office of Research and Development
U.S. Patent Office
Department of Commerce
Washington 25, D.C.

Andrews, J.C.
Technical Information Services
E.I. du Pont de Nemours & Co., Inc.
Savannah River Lab.
Aiken, S.C.

Angell, Richard S.
Subject Cataloging Div.
The Library of Congress
Washington 25, D.C.

Arendell, Frances H.
Warner-Lambert Research Institute
Tabor Rd.
Morris Plains, N.J.

Armstrong, Philip
Hughes Aircraft Co.
P.O. Box 2097
Fullerton, Calif.

Arnold, James M.
Recordak Corp.
415 Madison Ave.
New York, N.Y.

Ashour, Gladys B.
Emerson Research Labs.
1140 East-West Highway
Silver Spring, Md.

Astin, Allen V.
National Bureau of Standards
Washington 25, D.C.

Atwood, Wallace W., Jr.
Office of International Relations
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

Axford, S.J.
Vital Statistics Section, D.B.S.
Department of Trade and Commerce
Tunney's Pasture
Ottawa, Ontario, Canada

B

Babbitt, J.D.
Information Branch
National Research Council of Canada
100 Sussex Drive
Ottawa 2, Ontario, Canada

Baer, Karl A.
National Housing Center Library
1625 L St., N.W.
Washington 6, D.C.

Bagley, Philip R.
Lincoln Lab.
Massachusetts Institute of Technology
P.O. Box 73
Lexington 73, Mass.

Bahn, Gilbert S.
Marquardt Aircraft Co.
16555 Saticoy St.
Van Nuys, Calif.

Bailey, George D.
Administration Bldg.
Plant Industry Station
U.S. Department of Agriculture
Beltsville, Md.

Bailey, Lothrop H.
Gulf Research & Development Co.
P.O. Drawer 2038
Pittsburgh 30, Pa.

Baird, Norma
Computer Usage Co.
18 East 41st St.
New York 17, N.Y.

Baker, Dale B.
The Chemical Abstracts Service
The Ohio State University
Columbus 10, Ohio

Baker, W.O.
Murray Hill Lab.
Bell Telephone Labs., Inc.
Murray Hill, N.J.

Ball, Howard R.
Publications Branch, Bureau of Ships
Department of the Navy
Washington 25, D.C.

Ball, Norman T.
Interdepartmental Committee on Scientific Research & Development
National Science Foundation
1951 Constitution Ave., N.W.
Washington 25, D.C.

Ballard, Delbert L.
Documentation, Inc.
2521 Connecticut Ave., N.W.
Washington 8, D.C.

Baltes, Jack
Technical Library
Globe-Union, Inc.
900 East Keefe Ave.
Milwaukee 1, Wis.

Banes, Anthony
Computation Reduction Center & Language Translation
The Ramo-Wooldridge Corp.
P.O. Box 45534, Airport Station
Los Angeles 45, Calif.

Barden, William A.
Armed Services Technical Information Agency
Arlington Hall Station
Arlington 12, Va.

Bar-Hillel, Yehoshua
Faculty of Humanities
The Hebrew University
30 Abarbanel St.
Jerusalem, Israel

Barker, Dale L.
Price Gilbert Library
Georgia Institute of Technology
Atlanta 13, Ga.

Barrow, Valentine K.
Defence Research Board
Directorate of Scientific Information Services
125 Elgin St.
Ottawa, Ontario, Canada

Bartlett, Lynn J., Jr.
Information for Industry, Inc.
1000 Connecticut Ave., N.W.
Washington 6, D.C.

Barton, Henry A.
American Institute of Physics
335 East 45th St.
New York 17, N.Y.

Bateman, Betty B,
Office of the Librarian
Texas Engineers College
College Station, Tex.

Bateman, Mary Ruth
Boyce Thompson Institute for
Plant Research
1086 North Broadway
Yonkers 3, N.Y.

Bauer, Charles K.
Scientific and Technical Information Dept.
Lockheed Aircraft Corp.
86 South Cobb Drive
Marietta, Ga.

Baughman, Dorothy
Radiation Lab. Library
University of California
P.O. Box 808
Livermore, Calif.

Baumann, Dwight M.
Massachusetts Institute of Technology
Cambridge 38, Mass.

Baxendale, Phyllis
Research Lab.
I.B.M. Corp.
Monterey and Cottle Rds.
San Jose, Calif.

Bebbington, Marguerite
The International Nickel Co., Inc.
67 Wall St.
New York 5, N.Y.

Behnke, John A.
The Ronald Press Co.
15 East 26th St.
New York 10, N.Y.

Beinlich, J.J.
Research and Technology
U.S. Steel Corp.
525 William Penn Place
Pittsburgh 30, Pa.

Bejuki, Walter M.
Prevention of Deterioration Center
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

Benjamin, Curtis G.
McGraw-Hill Book Co., Inc.
330 West 42nd St.
New York 36, N.Y.

Benson, Bernard S.
Benson-Lehner Corp.
11930 West Olympic Blvd.
Los Angeles 64, Calif.

Benson, Frederic R.
Information Branch
Atlas Powder Co., Chemical Research Dept.
Wilmington 99, Del.

Benton, Mildred
U.S. Naval Research Lab.
Washington 25, D.C.

Berenato, Annamae
E.I. du Pont de Nemours & Co., Inc.
Eastern Lab.
Gibbstown, N.J.

Berend, Gertrude
Chemical Research Dept.
Hoffmann-La Roche, Inc.
Roche Park
Nutley 10, N.J.

Bernal, J.D.
Dept. of Physics
Birkbeck College
University of London
Malet St.
London W.C.1, England

Bernier, Charles L.
The Chemical Abstracts Service
The Ohio State University
Columbus 10, Ohio

Bertalan, Frank J.
U.S. Naval Research Lab.
Office of Naval Research
Washington 25, D.C.

Berthel, John H.
220 Gilman Hall
The Johns Hopkins University
Baltimore 18, Md.

Bigelow, Julian
School of Mathematics
The Institute for Advanced Study
Princeton, N.J.

Birnbaum, Henry
Office of the Director
National Bureau of Standards
Washington 25, D.C.

Bishop, Charles
University of Buffalo
Buffalo General Hospital
Buffalo 3, N.Y.

Black, Donald V.
Physics Library
University of California
Los Angeles 24, Calif.

Blair, Keith G.
Convair Library
Mail Zone 6-157
San Diego 12, Calif.

Blanchard, J.Richard
University of California Library
Davis, Calif.

Blanchard, Robert O.
I.B.M. Corp.
590 Madison Ave.
New York 22, N.Y.

Boardman, Brewer F.
Atomic Energy Div.
Phillips Petroleum Co.
Box 1259
Idaho Falls, Idaho

Boaz, Reynolds
11612 35th Ave.
Beltsville, Md.

Boeing, John K.
Recordak Corp.
415 Madison Ave.
New York, N.Y.

Bohnert, Lea M.
RCA Service Co.
838 North Henry St.
Alexandria, Va.

Boice, Theodore D.
E.I. du Pont de Nemours & Co., Inc.
Jackson Lab., P.O. Box 525
Wilmington 99, Del.

Bolton, Wallis D.
Data Processing Div.
I.B.M. Corp.
112 East Post Rd.
White Plains, N.Y.

Bonn, George S.
Graduate School of Library Service
Rutgers University
New Brunswick, N.J.

Bonnett, Howard T.
G.D.Searle & Co.
Box 5110
Chicago 80, Ill.

Bonniwell, Marion E.
Technical Library, Bureau of Ships
Department of the Navy
Washington 25, D.C.

Booth, Andrew D.
Dept. of Numerical Automation
University of London
Malet St.
London W.C.1, England

Boquet, Paul
Institut Pasteur
Annexe de Garches
28 rue du Docteur Roux
Paris, France

Borko, Harold
System Development Corp.
2500 Colorado Ave.
Santa Monica, Calif.

Bourgeois, Paul
I.C.S.U. Abstracting Board
Royal Observatory of Belgium
7 Avenue Circulaire
UCCLE Brussels 18, Belgium

Bourne, Charles
Stanford Research Institute Library
Menlo Park, Calif.

Bouthilet, Lorraine
Technical Information Unit
Psychopharmacology Service Center
National Institute of Mental Health
Bethesda 14, Md.

Boutry, G.A.
I.C.S.U. Abstracting Board
292 rue Saint-Martin
Paris 3, France

Boyles, John Harrison
Georgetown University Law Center Library
506 E St., N.W.
Washington 1, D.C.

Brady, James J.
AGT Library System
General Electric Co.
Cincinnati 15, Ohio

Brenner, Everett H.
The Dow Chemical Co.
James River Div. Library
Williamsburg, Va.

Bricker, J.L.
The Ramo-Wooldridge Corp.
P.O. Box 45215, Airport Station
Los Angeles 45, Calif.

Brisch, E.G.
Brisch, Inc.
1040 Secor Hotel Bldg.
Toledo 3, Ohio

Brode, Mildred Hooker
David Taylor Model Basin
Washington 7, D.C.

Brode, Robert B.
Office of the Director
National Science Foundation
1951 Constitution Ave., N.W.
Washington 25, D.C.

Brode, Wallace R.
Department of State
Washington 25, D.C.

Brodman, Estelle
National Library of Medicine
7th St. & Independence Ave., S.W.
Washington 25, D.C.

Broome, F.K.
Research Div.
Armour and Co.
1425 West 42nd St.
Chicago 9, Ill.

Brown, Curtis
The Institute of Paper Chemistry
Appleton, Wis.

Brown, Jack
National Research Council of Canada Library
100 Sussex Drive
Ottawa 2, Ontario, Canada

Brown, L.C.
Space Technology Labs.
P.O. Box 45564, Airport Station
Los Angeles 45, Calif.

Brown, Sir Lindor
The Royal Society
Burlington House, Piccadilly
London W.1, England

Browne, Secor D.
Massachusetts Institute of Technology Libraries
Cambridge 39, Mass.

Brownson, Helen L.
Office of Scientific Information
National Science Foundation
1951 Constitution Ave., N.W.
Washington 25, D.C.

Brunning, Dennis A.
Chester Beatty Research Inst.: Inst. of Cancer Research
Royal Cancer Hospital
Fulham Rd.
London S.W.3, England

Bry, Ilse
Editorial Committee
Mental Health Book Review Index
317 West 105th St.
New York 25, N.Y.

Bryant, Douglas W.
Harvard University Library
Cambridge 38, Mass.

Bryant, Margaret S.
U.S. Department of Agriculture Library
Washington 25, D.C.

Brygoo, Pierre
Institut Pasteur
28 rue du Docteur Roux
Paris 15, France

Bucans, Lawrence
3706 Randolph Rd.
Silver Spring, Md.

Buckland, Lawrence F.
Itek Corp.
1605 Trapelo Rd.
Waltham 54, Mass.

Bugher, John C.
The Rockefeller Foundation
49 West 49th St.
New York 20, N.Y.

Burlew, John S.
The Franklin Institute
20th & Benjamin Franklin Pkwy.
Philadelphia 3, Pa.

Burton, Robert Edward
Linde Co.
1500 Polco St.
Indianapolis 24, Ind.

Buscombe, William
The Australian National University
Box 4, G.P.O.
Canberra, A.C.T., Australia

Buzzati-Traverso, Adriano
Institute de Genetics
Universitià di Pavia
via Sant'Epifanio 14
Pavia, Italy

Byers, Dohrman H.
Occupational Health Information Exchange
Dept. of Health, Education, and Welfare
Bureau of State Services
1014 Broadway
Cincinnati 2, Ohio

C

Cagle, Fred R.
Dept. of Zoology
Tulane University
New Orleans 18, La.

Cahoon, Lynda Lee
Div. of Research Grants
National Institutes of Health
303 Colemont Bldg.
Silver Spring, Md.

Caianiello, Eduardo R.
Institute for Theoretical and Nuclear Physics
Padiglione 19, Mostra d'Oltremare
Naples, Italy

Callanan, Joan
Office of Critical Tables
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

Calvano, Mary F.
Technical Library
American Sterilizer Co.
2424 West 23rd St.
Erie, Pa.

Campbell, Charles I.
The Rockefeller Institute
66th St. and York Ave.
New York 21, N.Y.

Campbell, H.C.
The Toronto Public Library
College and St. George Sts.
Toronto 2B, Ontario, Canada

Campbell, H.F.
The Procter & Gamble Co.
M.A. & R. Bldg., Ivorydale
Cincinnati 17, Ohio

Campbell, Wayne R.
6135 35th St., North
Arlington 13, Va.

Canova, Madeline Frances
U.S. Naval Ordnance Test Station
Box 674
China Lake, Calif.

Canter, Louis
Convair-Astronautics Library
P.O. Box 1128
San Diego 12, Calif.

Caponio, Virginia Hall
Office of Scientific Information
National Science Foundation
1951 Constitution Ave., N.W.
Washington 25, D.C.

Carmichael, Leonard
Smithsonian Institution
Washington 25, D.C.

Carmichael, Robert L.
Bell Telephone Labs., Inc.
Whippany, N.J.

Carnes, Kathleen
Technical Information Section
Watertown Arsenal
Watertown 72, Mass.

Carson, Eileen R.
Div. of Bldg. Research
National Research Council of Canada
Montreal Rd.
Ottawa 2, Ontario, Canada

Casagandra, Elizabeth A
Research Corp.
405 Lexington Ave.
New York 17, N.Y.

Casale, Ina B.
E.I. du Pont de Nemours & Co., Inc.
Engineering Research Lab., Experimental Station
Wilmington 98, Del.

Case, Charles Z.
Eastman Kodak Co.
343 State St.
Rochester 4, N.Y.

Cattaneo, A.
Fiat Co.
Corso Giovanni Agnelli 200
Torino, Italy

Chamberlin, Waldo
Graduate School of Arts and Science
New York University
Washington Square
New York 3, N.Y.

Chandler, James G.
Woodlawn Manor
601 Stillwell Ave.
Alexandria, Va.

Chaplin, C.E.
Entomology Div.
Science Service of Canada
Science Service Bldg.
Ottawa, Ontario, Canada

Cherenin, V.P.
All-Union Institute of Scientific and Technical Information
Academy of Sciences of the U.S.S.R.
Baltiiskii Poselok 42-B
Moscow 57, Russia

Cherni, Arkadi Ivanovich
All-Union Institute of Scientific and Technical Information
Academy of Sciences of the U.S.S.R.
Baltiiskii Poselok 42-B
Moscow 57, Russia

Chesnut, Charlotte F.
Technical Documents Library
Army Ballistic Missile Agency
Mims Court G-3
Huntsville, Ala.

Chilton, L.V.
Intelligence Dept.
The British Aluminium Co., Ltd.
Norfolk House, St. James's Square
London S.W.1, England

Citron, Joan
System Development Corp.
2400 Colorado Ave.
Santa Monica, Calif.

Clapp, Verner W.
Council on Library Resources, Inc.
Suite 715, 1025 Connecticut Ave., N.W.
Washington 6, D.C.

Claridge, P.R.P.
Low Temperature Research Station
Cambridge, England

Clark, Harold E.
Haloid Xerox, Inc.
2 Haloid St.
Rochester 3, N.Y.

Clark, Joe H.
Technical Information Services
Pearl River Labs., American Cyanamid Co.
Middletown Rd.
Pearl River, N.Y.

Clark, Mabel D.
Technical Reference Library
Naval Medical Research Institute
Bethesda 14, Md.

Clark, Noble
College of Agriculture
University of Wisconsin
Madison 6, Wis.

Clark, Ralph L.
Washington Office
Stanford Research Institute
711 14th St., N.W.
Washington 5, D.C.

Clausing, Pieter
Philips Research Labs.
Eindhoven, Netherlands

Cleland, Reina G.
Solid State Div., Oak Ridge National Lab.
P.O. Box X
Oak Ridge, Tenn.

Cleverdon, Cyril W.
College of Aeronautics
Cranfield, Bletchley
Bucks, England

Cloud, Gould H.
Jersey Production Research Co.
1133 North Lewis Ave.
Tulsa, Okla.

Coe, Conway P.
Radio Corp. of America Patent Dept.
711 14th St., N.W.
Washington, D.C.

Coffman, Joseph W.
U.S. Army Ordnance Missile Command
Army Ballistic Missile Agency
Redstone Arsenal, Ala.

Coghlan, Owen, Jr.
Alphabib Co.
3237 Griest Ave.
Cincinnati 8, Ohio

Cohen, Aaron
Building 5–132
General Electric Co.
1 River Rd.
Schenectady 5, N.Y.

Cohen, Marvin
Astro-Electronic Products Div.
Radio Corp. of America
Princeton, N.J.

Cole, Peter Frank
The British Petroleum Co., Ltd.
Britannic House, Finsbury Circus
London E.C.2, England

Coleman, Earl M.
Consultants Bureau, Inc.
227 West 17th St.
New York 11, N.Y.

Collier, Charles H.
Engineering Dept.
E.I. du Pont de Nemours & Co., Inc.
Wilmington 98, Del.

Comstock, Marjorie
Research Library, Brookhaven National Lab.
Associated Universities, Inc.
Upton, Long Island, N.Y.

Connelly, John W., Jr.
111 West Hunting Towers
Alexandria, Va.

Conrad, C.C.
Textile Fibers Dept.
E.I. du Pont de Nemours & Co., Inc.
902 Orange St.
Wilmington 98, Del.

Conrad, G.Miles
Biological Abstracts
3815 Walnut St.
Philadelphia 4, Pa.

Cook, John S.
Coast and Geodetic Survey
U.S. Department of Commerce
Washington 25, D.C.

Cooke, Charles C.
Technical Center
Owens-Illinois Glass Co.
P.O. Box 1035
Toledo 1, Ohio

Coolidge, Harold J.
Pacific Science Board
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

Cooper, Franklin S.
Haskins Labs., Inc.
305 East 43rd St.
New York 17, N.Y.

Cordonnier, Gerard A.
1 rue de la Ville l'Evegne
Paris 8, France

Cornelius, Merlin E.
1911 Laguna Rd.
Adelphi, Md.

Cortelyou, Ethaline H.
Illinois Inst. of Technology
Armour Research Foundation
10 West 35th St.
Chicago 16, Ill.

Coupland, Richard C., Jr.
European Office, ARDC
Shell Bldg.
47 rue Cantersteen
Brussels, Belgium

Courtis, Reginald P.
Cataloging Div.
Armed Forces Supply Support Center
19th and E. Capitol Sts., N.E.
Washington 25, D.C.

Covell, Louis C.
R.F.D. 1, Box 646
Fairfax, Va.

Cowgill, Logan O.
Office, Chief of Engineers, Library Br.
U.S. Department of the Army
Washington 25, D.C.

Craig, Robert L.
McGraw-Hill Book Co., Inc.
330 West 42nd St.
New York 36, N.Y.

Crandall, George S.
Research Dept.
Socony Mobil Oil Co.
Paulsboro, N.J.

Crane, E.J.
The Chemical Abstracts Service
The Ohio State University
Columbus 10, Ohio

Crane, Helen G.
American Chemical Society
1788 Coventry Rd.
Columbus 12, Ohio

Craven, Ruby S.
Aerodynamics Lab.
David Taylor Model Basin
Washington 25, D.C.

Crawford, Susan Y.
American Medical Assoc.
535 North Dearborn St.
Chicago 10, Ill.

Crowther, B.M.
Science Abstracts
The Institution of Electrical Engineers
Savoy Place
London W.C.2, England

Cummins, J.E.
Div. of Scientific and Technical Information
International Atomic Energy Agency
Vienna 1, Kaerntnerring, Austria

Cuthbert, F.Leicester
National Lead Co. of Ohio
P.O. Box 158, Mount Healthy Station
Cincinnati 31, Ohio

Cutlip, Mary E.
Research and Development Dept.
Naval Powder Factory
Indian Head, Md.

D

Dabrowski, Kazimierz
Polish Academy of Sciences
Palac Kultury I Nauki
Warsaw, Poland

Daniels, Marietta
Columbus Memorial Library
Pan American Union
1700 Constitution Ave., N.W.
Washington 6, D.C.

David, Charles W.
Longwood Library
Kennett Square, Pa.

Davis, Lora-Frances
M.D.Anderson Hospital & Tumor Inst.
The University of Texas
Houston 25, Tex.

Davison, Margaret C.
Bureau of Medicine, New Drug Branch
Department of Health, Education, and Welfare
Washington 25, D.C.

Dawson, Roy C.
Food and Agriculture Organization of the United Nations
1535 C St., S.W.
Washington 25, D.C.

Day, Melvin S.
Div. of Information Services
U.S. Atomic Energy Commission
Washington 25, D.C.

Dean, Crowell O.
Technical Information Center
A.E.Staley Mfg. Co.
Decatur, Ill.

de Grolier, Eric
Centre Français d'Échanges et de Documentation Techniques
32, Corso Magenta
Milan, Italy

deHaan, C.J.
Netherlands Patent Office
6 Willem Witsenplein
The Hague, Netherlands

Del Frate, Adelaide A.
Research and Development Div.
W.R.Grace & Co.
Washington Research Center
Clarksville, Md.

Dellinger, J.Howard
3900 Connecticut Ave., N.W.
Washington 8, D.C.

De Lucia, Al
Rome Air Development Center
Griffiss Air Force Base
Rome, N.Y.

Denison, J.T.
Polychemicals Dept.
E.I. du Pont de Nemours & Co., Inc.
Wilmington 98, Del.

Dennis, Bernard K.
Central Library
General Electric Co.
Cincinnati 15, Ohio

Depelsenaire, Robert
Laboratoires de Recherches Purfina S.A.
Chaussée de Vilvorde 100
Brussels 12, Belgium

Dere, John R.
Research Dept.
I.B.M. Corp.
P.O. Box 3529, Rincon Annex
San Francisco, Calif.

de Stevens, George
Ciba Pharmaceutical Products, Inc.
556 Morris Ave.
Summit, N.J.

de Tonnancour, P.R.
Convair Library
5908 Lowell Ave.
Fort Worth 7, Tex.

Devlin, T.J.
Technical Information Div.
Esso Research & Engineering Co.
P.O. Box 51
Linden, N.J.

Dickison, Raymond R.
Oak Ridge National Lab. Library
P.O. Box X
Oak Ridge, Tenn.

Diels, Jan C.
Netherlands Embassy
1470 Euclid St., N.W.
Washington 9, D.C.

Dillon, Robert M.
Educational Facilities Labs., Inc.
447 Madison Ave.
New York, N.Y.

Dittberner, DonaldL.
I.B.M. Corp.
1111 Connecticut Ave., N.W.
Washington 6, D.C.

Dodd, James B.
Atomic Energy Div.
The Babcock and Wilcox Co.
P.O. Box 1260
Lynchburg, Va.

Dodge, P.O.
Computer Dept.
General Electric Co.
Phoenix, Ariz.

Dodge, William
131 Springvale Ave.
McLean, Va.

Donley, Albert M., Jr.
Northeastern University Library
360 Huntington Ave.
Boston 15, Mass.

Doran, Margaret T.
Dept. of Medicine and Surgery
Veterans Administration
Washington 25, D.C.

Dorn, Harold F.
Div. of Research Services
Biometrics Branch
National Institutes of Health
Bethesda 14, Md.

Doss, Mildred A.
Animal Disease and Parasite Research Branch
Agricultural Research Service
U.S. Department of Agriculture
Beltsville, Md.

Dostert, Leon E.
Institute of Language and Linguistics
Georgetown University
Washington 7, D.C.

Dowling, E.C.
Sandia Corp.
Sandia Base
Albuquerque, N. Mex.

Doyle, Lauren B.
System Development Corp.
2400 Colorado St.
Santa Monica, Calif.

Draper, Wesley
Medical Society of the County of Kings
1313 Bedford Ave.
Brooklyn 16, N.Y.

Drenowatz, Margaret C.
Tobacco Literature Service
7 D.H. Hill Library
North Carolina State College
Raleigh, N.C.

Drescher, Milton A.
Science and Industry Dept.
Milwaukee Public Library
814 West Wisconsin Ave.
Milwaukee 3, Wis.

Dubester, Henry J.
Reference Dept.
The Library of Congress
Washington 25, D.C.

Duggan, Maryann
Magnolia Petroleum Co.
P.O. Box 900
Dallas 21, Tex.

Duncombe, Charles G.
Research Institute
University of Detroit
4001 West McNichols Rd.
Detroit 21, Mich.

Du Shane, Graham
American Assoc. for the Advancement of Science
1515 Massachusetts Ave., N.W.
Washington 5, D.C.

Duval, Raymonde
Bulletin Signaletique
Centre National de la Recherche Scientifique
15 Quai Anatole
Paris, France

Dyke, Freeman H.
Engineering Dept.
E.I. du Pont de Nemours & Co., Inc.
Wilmington 98, Del.

Dyke, H.Gordon
International Latex Corp.
Playtex Park
Dover, Del.

Dyson, George Malcolm
1 Burton Walks
Loughborough, Leics., England

E

Eastman, William
The Baltimore & Ohio Railroad Co.
2 North Charles St.
Baltimore 1, Md.

Eaton, Elizabeth S.
Research Labs.
Mead Johnson & Co.
Evansville 21, Ind.

Ebert, H.
Physikalisch-Technische Bund.
Bundesallee 100, (20b)
Braunschweig, Germany

Edlund, Margaret F.
Information Div., Radiation Lab.
University of California
Berkeley 4, Calif.

Edmundson, Harold P.
Engineering Div.
The Rand Corp.
1700 South Main St.
Santa Monica, Calif.

Egerton, Sir Alfred
Dept. of Chemical Engineering
Imperial College
London S.W. 7, England

Eisele, Harold F.
U.S. Public Health Service
Department of Health, Education, and Welfare
Washington 25, D.C.

Eisenmann, Jacqueline
Centre National de la Recherche Scientifique
Centre de Documentation
16 rue Pierre-Curie
Paris 5, France

Elion, Edward
Technical Representations Co.
1622 Juniper St., N.W.
Washington 12, D.C.

Ellinger, Werner B.
American Assoc. of Law Libraries
The Library of Congress
Washington 25, D.C.

Elliott, Owen
Butterworth Publishers
London, England

Elting, E.C.
Agricultural Research Service
U.S. Department of Agriculture
Washington 25, D.C.

Embrey, Lee Anna
Office of the Director
National Science Foundation
1951 Constitution Ave., N.W.
Washington 25, D.C.

Emerson, Mary E.
Institute of Textile Technology
Charlottesville, Va.

Emerson, William L.
Science-Industry Dept.
Long Beach Public Library
Long Beach 2, Calif.

Erlandson, Paul M.
Central Research & Engineering Div.
Continental Can Co., Inc.
7622 South Racine Ave.
Chicago 20, Ill.

Erskine, Hugh S.
1128 Panley Rd.
Silver Spring, Md.

Escande, Leopold
Academy of Sciences of Paris
4 Blvd. Riquet
Toulouse (Hte. Garonne), France

Esterquest, Ralph T.
Harvard Medical School Library
25 Shattuck St.
Boston 15, Mass.

Estrin, Gerald
Dept. of Engineering
University of California
Los Angeles 24, Calif.

Estrin, Thelma A.
12414 Dearbrook Lane
Los Angeles 49, Calif.

F

Faden, Ben
Watson Lab.
Data Processing Div.
I.B.M. Corp.
612 West 116th St.
New York 25, N.Y.

Fairthorne, R.A.
Ministry of Supply
Royal Aircraft Establishment
South Farnborough
Hants, England

Fano, Robert M.
Research Lab. of Electronics
Massachusetts Institute of Technology
Cambridge 39, Mass.

Farley, Richard A.
University of Nebraska Libraries
Lincoln 8, Nebr.

Farradane, J.E.L.
Tate & Lyle Ltd., Research Labs.
Keston, Kent, England

Feldman, William
Apparatus and Optical Div.
Eastman Kodak Co.
400 Plymouth Ave., North
Rochester 4, N.Y.

Fell, Norman
Hazelton Labs., Inc.
Box 333
Falls Church, Va.

Felter, Jacqueline W.
Lee Coombe Memorial Library
Memorial Center for Cancer and Allied Diseases
444 East 68th St.
New York 21, N.Y.

Ferguson, Malcolm S.
Technical Reports and Library
Communicable Disease Center
U.S. Public Health Service
50 Seventh St., N.E.
Atlanta 23, Ga.

Fink, Dan T.
The Danish National Centre for Building Documentation
3 Tordenskjoldsgade
Copenhagen, Denmark

Fiordalisi, Vincent E.
Rutgers University Law School
53 Washington St.
Newark 2, N.J.

Firth, Margaret A.
Research Div.
United Shoe Machinery Corp.
Balch St.
Beverly, Mass.

Fishenden, R.M.
Research Group
United Kingdom Atomic Energy Authority
Harwell, Didcot
Berks, England

Fisher, Norman G.
E.I. du Pont de Nemours & Co., Inc.
Experimental Station
Wilmington 98, Del.

Fitzgerald, Edward L.
Remington Rand Univac
Sperry Rand Corp.
315 4th Ave.
New York 10, N.Y.

Flanagan, J.
Data Processing Div.
I.B.M. Corp.
112 East Post Rd.
White Plains, N.Y.

Fleischer, Michael
U.S. Geological Survey
Department of the Interior
Washington 25, D.C.

Fleischmann, Isaac
U.S. Patent Office
Department of Commerce
Washington 25, D.C.

Fleming, Thomas P.
Columbia University
630 West 168th St.
New York 32, N.Y.

Focke, Helen M.
School of Library Science
Western Reserve University
Cleveland 6, Ohio

Fomenko, Sergei M.
Control Systems Div.
The Ramo-Wooldridge Corp.
P.O. Box 45534, Airport Station
Los Angeles 45, Calif.

Fonseca, Edson Nery da
Brazilian Assoc. of Librarians
Rua Pedro Lessa 36, 2° andar
Rio de Janiero, D.F., Brazil

Ford, Robert
Pharmacology Div.
Institute for Therapeutic Research
Merck Sharp & Dohme Research Labs.
West Point, Pa.

Foster, Baba
Operations Evaluation Group
Office of the Chief of Naval Operations
Department of the Navy
Washington 25, D.C.

Foulke, Jean E.
The New York Academy of Medicine
2 East 103rd St.
New York 29, N.Y.

Fourdrinier, Sylvia
American Institute of Chemical Engineers
25 West 45th St.
New York 36, N.Y.

Foust, W.
The Computation Lab. of Harvard University
Cambridge 38, Mass.

Frankel, Joan
Newton College of the Sacred Heart
885 Centre St.
Newton 59, Mass.

Frauendorfer, Sigmund V.
Oesterreichische Gesellschaftfür Dokumentation und Bibliographie
Gregor Mendelstr. 33
Vienna 1, Austria

Freund, Marjorie M.
1421 Massachusetts Ave., N.W.
Washington, D.C.

Fried, Ferdinand
Ayerst, McKenna and Harrison, Ltd.
P.O. Box 6115
Montreal, Quebec, Canada

Friedlander, M.O.
Engineering Library, Plant 5
Grumman Aircraft Engineering Corp.
Bethpage, Long Island, N.Y.

Friedman, Corrine
Morris D.Friedman, Inc.
67 Reservoir St.
Needham Heights 94, Mass.

Friedman, Herbert J.
The Chemical Abstracts Service
The Ohio State University
Columbus 10, Ohio

Friedrich, William
Assoc. for International Development
2122 Massachusetts Ave., N.W.
Washington 8, D.C.

Frome, Julius
Office of Research and Development
U.S. Patent Office
Department of Commerce
Washington 25, D.C.

Frommherz, Carl J.
Chicago Undergraduate Div.
University of Illinois Library
Navy Pier
Chicago 11, Ill.

Fry, Bernard M.
Office of Scientific Information
National Science Foundation
1951 Constitution Ave., N.W.
Washington 25, D.C.

Fuller, Margaret H.
American Iron and Steel Institute
150 East 42nd St.
New York 16, N.Y.

Fusslerl, Herman H.
The University of Chicago Library
1116 East 59th St.
Chicago 37, Ill.

G

Gaarder, Sydney R.
Office of Industrial Development
U.S. Atomic Energy Commission
Washington 25, D.C.

Gabel, M.W.
Eastman Kodak Co.
343 State St.
Rochester 3, N.Y.

Gabriel, Henry
Technical Information Section
Research and Development Dept.
Phillips Petroleum Co.
Bartlesville, Okla.

Gamble, Dean F.
Cancer Chemotherapy National Service Center
National Institutes of Health
Bethesda 14, Md.

Gardin, J.C.
Centre Nationale de la Recherche Scientifique
32 rue Vancau
Paris 7, France

Gardner, John L.
Research Library
Allen B. du Mont Labs., Inc.
35 Market St.
East Patterson, N.J.

Gardner, Lytt I.
Dept. of Pediatrics
Upstate Medical Center
State University of New York
766 Irving Ave.
Syracuse 10, N.Y.

Garfield, Eugene E.
Eugene Garfield Associates
1523 Spring Garden St.
Philadelphia 30, Pa.

Gass, Saul I.
I.B.M. Corp.
1111 Connecticut Ave., N.W.
Washington 6, D.C.

Gates, Marshall
Dept. of Chemistry
University of Rochester
River Campus Station
Rochester 20, N.Y.

Geer, Harriet A.
Research Div.
Parke, Davis & Co.
Joseph Campau at the River
Detroit 32, Mich.

Gibbs, R.Clifton
Nuclear Data Project
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

Gibson, Eleanor B.
Logan Lewis Library
Carrier Corp.
Carrier Pkwy.
Syracuse 1, N.Y.

Gibson, R.E.
Applied Physics Lab.
The Johns Hopkins University
8621 Georgia Ave.
Silver Spring, Md.

Gibson, Robert W., Jr.
Information Management Div.
Battelle Memorial Institute
505 King Ave.
Columbus 1, Ohio

Ginsberg, Helen F.
Schering Corp.
60 Orange St.
Bloomfield, N.J.

Giuliano, Vincent E.
The Computation Lab. of Harvard University
Cambridge 38, Mass.

Gladding, Elinor H.
E.I. du Pont de Nemours & Co., Inc.
Jackson Lab.
P.O. Box 1407
Wilmington 98, Del.

Glass, H.Bentley
Dept. of Biology
The Johns Hopkins University
Baltimore 18, Md.

Goldstein, Gordon D.
Information Systems Branch
Office of Naval Research
Washington 25, D.C.

Good, I.J.
25 Scott House
Princess Elizabeth Way
Cheltenham, Gloustershire, England

Gosnell, Charles Francis
State Education Dept.
The New York State Library
University of the State of New York
Albany 1, N.Y.

Goss, Dorothy K.
Control Systems Div.
The Ramo-Wooldridge Corp.
P.O. Box 45534, Airport Station
Los Angeles 45, Calif.

Gosset, Marion
United Kingdom Atomic Energy Authority
Research Group
Atomic Energy Research Establishment
Harwell, Didcot
Berks, England

Gould, S.H.
Mathematical Reviews
190 Hope St.
Providence 6, R.I.

Gove, N.B.
Nuclear Data Project
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

Gove, Philip B.
G. & C. Merriam Co.
Federal St.
Springfield 2, Mass.

Gray, Dwight E.
Office of Scientific Information
National Science Foundation
1951 Constitution Ave., N.W.
Washington 25, D.C.

Green, John
Office of Technical Services
Department of Commerce
Washington 25, D.C.

Green, Myrtle
Div. of Applied Chemistry
National Research Council of Canada
Montreal Rd.
Ottawa, Ontario, Canada

Greene, Charles E.
Div. of Research Services
National Institutes of Health
Bethesda 14, Md.

Greenwell, B.E.
Abbott Labs.
North Chicago, Ill.

Gregory, E.F.
Special Engineering Product Div.
I.B.M. Corp.
425 Park Ave.
New York, N.Y.

Grell, Norbert F.
International Atomic Energy Agency
Vienna, 1, Kaerntnerring, Austria

Grenier, Myra T.
Aerojet-General Corp. Library
P.O. Box 296
Azusa, Calif.

Gresham, William A., Jr.
Engineering Experiment Station
Georgia Institute of Technology
Atlanta 13, Ga.

Gribbin, John H.
Library
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

Griffin, Marjorie
I.B.M. Corp. Library
Monterey and Cottle Rds.
San Jose, Calif.

Griffith, John
I.B.M. Research Center
P.O. Box 390
Poughkeepsie, N.Y.

Groven, Louis
Belgium Embassy
3330 Garfield St., N.W.
Washington 8, D.C.

Gull, C.D.
Computer Dept.
General Electric Co.
3240 Wilson Blvd.
Arlington, Va.

Gurk, Herbert
Astro-Electronic Products Div.
Radio Corp. of America
Princeton, N.J.

Gvozdev, Yuri I.
Embassy of the U.S.S.R.
1706 16th St., N.W.
Washington, D.C.

H

Hagen, Peter John
Naval Research Lab.
Department of the Navy
Washington 25, D.C.

Haglind, J.B.
Olin Mathieson Chemical Co.
P.O. Box 480
Niagara Falls, N.Y.

Halbert, Michael H.
Operations Research Group
Case Institute of Technology
Cleveland 6, Ohio

Hall, Ernest E.
Industry Assistance Br.
U.S. Atomic Energy Commission
Washington 25, D.C.

Hall, Joseph E.
Air Information Div.
The Library of Congress
Washington 25, D.C.

Halpern, George
The Martin Co.
Mail J398, Middle River
Baltimore, Md.

Hamrick, Lillian
Office of Technical Services
Department of Commerce
Washington 25, D.C.

Handrick, G.R.
Arthur D.Little, Inc.
15 Acorn Park
Cambridge 40, Mass.

Harman, R.Joyce
Lincoln Lab.
Massachusetts Institute of Technology
P.O. Box 73
Lexington 73, Mass.

Harper, Lawrence A.
Dept. of History
University of California
Berkeley 4, Calif.

Harris, Zellig
Dept. of Linguistics
University of Pennsylvania
Philadelphia 4, Pa.

Harrison, Marvin
Research and Development Dept.
Eastman Kodak Co.
400 Plymouth Ave., North
Rochester 4, N.Y.

Harte Robert A.
Merck Sharp & Dohme Research Labs.
West Point, Pa.

Hartman, Frank W.
Office of the Surgeon General
U.S. Air Force
Washington 25, D.C.

Hartwell, Jonathan L.
National Cancer Institute
National Institutes of Health
Bethesda 14, Md.

Hattery, Lowell H.
American University
1901 F St., N.W.
Washington 6, D.C.

Hayes, Robert M.
The Magnavox Co.
2255 Carmelina Ave.
Los Angeles 64, Calif.

Hayne, Robert L.
Smith, Kline & French Labs.
1530 Spring Garden St.
Philadelphia 1, Pa.

Hazel, F.P.
Technical Information Dept.
Mitre Corp.
Box 244
Lexington, Mass.

Heald, J.Heston
Armed Services Technical Information Agency
Arlington Hall Station
Arlington 12, Va.

Heatherington, Daniel
Smith, Kline & French Labs.
1500 Spring Garden St.
Philadelphia 1, Pa.

Heatwole, Thelma C.
1411 Churchville Ave.
Staunton, Va.

Hedenius, Per J.
Swedish Society of Medical Physics and Techniques
34 Barnängsgatan
Stockholm Sö, Sweden

Heilprin, Laurence B.
Council on Library Resources, Inc.
1025 Connecticut Ave., N.W.
Washington 6, D.C.

Hellwege, Anne Marie
Institute of Technical Physics
Darmstadt, West Germany

Hellwege, Karl H.
Institute of Technical Physics
Darmstadt, West Germany

Henderson, Madeline Berry
Office of Scientific Information
National Science Foundation
1951 Constitution Ave., N.W.
Washington 25, D.C.

Henkle, Herman H.
The John Crerar Library
86 East Randolph St.
Chicago 1, Ill.

Henninger, Richard
Hazelton Labs., Inc.
Box 333
Falls Church, Va.

Henshaw, Paul S.
Div. of Biology and Medicine
U.S. Atomic Energy Commission
Washington 25, D.C.

Herdegen, Theodore F.
Smith, Kline & French Labs.
1530 Spring Garden St.
Philadelphia 1, Pa.

Herner, Mary
Herner and Co.
2625 Eye St., N.W.
Washington 7, D.C.

Herner, Saul
Herner and Co.
2625 Eye St., N.W.
Washington 7, D.C.

Herscher, Marvin B.
Radio Corp. of America
Camden 2, N.J.

Hertz, David Bendel
Arthur Andersen & Co.
67 Broad St.
New York 4, N.Y.

Herzbrun, Philip Ingram
Applications Engineering Section
National Bureau of Standards
Washington 25, D.C.

Heumann, Karl F.
The Chemical Abstracts Service
The Ohio State University
Columbus 10, Ohio

Heyliger, George E.
Control Systems Div.
The Ramo-Wooldridge Corp.
P.O. Box 45534, Airport Station
Los Angeles 45, Calif.

Hilsenrath, Joseph
Thermodynamics Section, Heat Div.
National Bureau of Standards
Washington 25, D.C.

Himwich, Williamina A.
Galesburg State Research Hospital
Galesburg, Ill.

Hirsch, Alice
Operations Research Office
The Johns Hopkins University
6935 Arlington Rd.
Bethesda, Md.

Hiscocks, E.Stanley
United Kingdom Scientific Mission
1907 K St., N.W.
Washington 6, D.C.

Hlobil, J.S.J.
General Research Co.
407 Commercial Center St.
Beverly Hills, Calif.

Ho, Don T.
Grace Research and Development Div.
W.R.Grace & Co.
Washington Research Center
Clarksville, Md.

Hodges, Mark W.
United Kingdom Scientific Mission
1907 K St., N.W.
Washington 6, D.C.

Hodgson, James G.
Library Branch
Quartermaster Food and Container Inst.
1819 West Pershing Rd.
Chicago 9, Ill.

Hodgson, Magdalene F.
American Medical Assoc.
535 North Dearborn St.
Chicago 10, Ill.

Hoegberg, Erick I.
Research and Development Div.
American Viscose Corp.
Marcus Hook, Pa.

Hoffman, Arthur C.
American Psychological Assoc.
1333 16th St., N.W.
Washington 6, D.C.

Hogg, I.H.
United Kingdom Atomic Energy Authority
Research and Development Br.
Risley, Warrington
Lancs, England

Holleman, W.Roy
Scripps Institution of Oceanography Library
University of California
La Jolla, Calif.

Holz, F.M.
1100 25th St., N.W.
Washington, D.C.

Holzapfel, Ruth L.
Technical Library
Silicone Products Dept.
General Electric Co.
Waterford, N.Y.

Honaman, R.Karl
Publications
Bell Telephone Labs, Inc.
463 West St.
New York, N.Y.

Hooker, Marjorie
Geological Survey
U.S. Department of the Interior
Washington 25, D.C.

Hooker, Ruth H.
Naval Research Lab.
Department of the Navy
Washington 25, D.C.

Hopp, Ralph H.
University of Minnesota Libraries
Minneapolis 14, Minn.

Hopper, Jean G.
Business, Science & Industry Dept.
Free Library of Philadelphia
Logan Square
Philadelphia 3, Pa.

Hoseh, M.
Chemical Abstracts
The Library of Congress
Washington 25, D.C.

House, Warren
RR #3
Herndon, Va.

Houtchens, Jean H.
Bldg. Research Institute
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

Houwink, R.
EURATOM
51-53 rue Belliard
Brussels, Belgium

Howard, Sir Herbert
Commonwealth Agricultural Bureaux
Farnham Royal
Bucks, England

Howe, Mary Taylor
Decatur Public Library
457 North Main St.
Decatur, Ill.

Howerton, Paul W.
3807 Everett St.
Kensington, Md.

Hudson, P.S.
Commonwealth Bureau of Plant Breeding and Genetics
Downing St.
Cambridge, England

Humphrey, Judith M.
Animal Parasite Labs.
U.S. Department of Agriculture
Beltsville, Md.

Hunt, Edna F.
National Research Council of Canada
100 Sussex Drive
Ottawa 2, Ontario, Canada

Hutchinson, Joan M.
National Lead Co. of Ohio Library
P.O. Box 158, Mt. Healthy Station
Cincinnati 31, Ohio

Hutchisson, Elmer
American Institute of Physics
335 East 45th St.
New York 17, N.Y.

Hyer, Anna L.
Dept. of Audio-Visual Instruction
National Education Assoc.
1201 16th St., N.W.
Washington 6, D.C.

Hyslop, Marjorie R.
American Society for Metals
7301 Euclid Ave.
Cleveland 3, Ohio

I

Igelsrud, Iver
Battelle Memorial Institute
505 King Ave.
Columbus 1, Ohio

Inverarity, Robert Bruce
Route #3
Ridgefield, Conn.

Irving, Wilbur C.
Operations Research Group
Geological Survey
U.S. Department of the Interior
Washington 25, D.C.

Irwin, Ralph R.
Navy Management Office
Department of the Navy
Washington 25, D.C.

Isakovic, Dusan J.
Yugoslav Center for Technical & Scientific Documentation
16 Admirala Geprata
Belgrade, Yugoslavia

Iung, Jean Pierre
Centre d'Études Nucleaires de Saclay
Boìte Postale 2, Gif-sur-Yvette
Seine et Oise, France

J

Jackson, Eugene B.
Research Staff Library
General Motors Corp. Library
P.O. Box 188, North End Station
Detroit 2, Mich.

Jackson, John B.
I.B.M. Corp.
1111 Connecticut Ave., N.W.
Washington 6, D.C.

Jahoda, Gerald
Esso Research & Engineering Co.
P.O. Box 51
Linden, N.J.

James, Peter
I.B.M. Research Center
Box 218
Yorktown Heights, N.Y.

Jaworski, Louis J.
Borden Chemical Co.
350 Madison Ave.
New York 17, N.Y.

Jenkins, Frances Briggs
University of Illinois Library
Urbana, Ill.

Jepson, Joy N.
Operations Evaluation Group
Office, Chief of Naval Operations
Washington 25, D.C.

Jessel, Walter
Route #3, Box 129
Vienna, Va.

Johnson, Arthur B.
California Research Corp.
P.O. Box 1627
Richmond 1, Calif.

Johnson, Carl M.
Technical Information Div.
U.S. Navy Electronics Lab.
San Diego 52, Calif.

Johnson, Elizabeth W.
Philadelphia College of Pharmacy and Medicine Library
Philadelphia 4, Pa.

Johnson, Irma Y.
Charles Hayden Library
Massachusetts Institute of Technology
Cambridge 38, Mass.

Johnson, Kathrine E.
Electronic Defense Lab.
Sylvania Electric Products, Inc.
P.O. Box 205
Mountain View, Calif.

Johnson, Quentin S.
Systems Development Div.
Benson-Lehner Corp.
11930 West Olympic Blvd.
Los Angeles 64, Calif.

Johnson, Walter J.
Academic Press, Inc.
111 Fifth Ave.
New York 3, N.Y.

Jolliff, Wade A., Jr.
Office of Long Range Requirements
U.S. Air Force Security Service
San Antonio, Tex.

Jones, Dorothy
Georgia Institute of Technology Library
Atlanta 13, Ga.

Jones, Lawrence Francis
Westinghouse Electric Corp.
P.O. Box 746
Baltimore 3, Md.

Jones, William S.
Computing Lab. 82-338
Union Carbide Chemicals Co.
P.O. Box 8004
South Charleston 3, W.Va.

Jonker, Frederick
Jonker Business Machines, Inc.
3939 Legation St., N.W.
Washington 15, D.C.

Jordan, Lloyd
1660 Lanier Place, N.W.
Washington, D.C.

Jorgenson, William E.
U.S. Navy Electronics Lab.
San Diego 52, Calif.

Joseph, G.S.
Burmah-Shell
Esplanade: Madras 1
South India

Josselson, Harry
Computation Lab., College of Liberal Arts
Wayne State University
Detroit 2, Mich.

Joyce, J.W.
Office of Special International Programs
National Science Foundation
1951 Constitution Ave., N.W.
Washington 25, D.C.

Judkins, Esther
The Rockefeller Institute Library
66th St. and York Ave.
New York 21, N.Y.

Juhasz, Stephen
Applied Mechanics Reviews
Southwest Research Institute
8500 Culebra Rd.
San Antonio, Tex.

K

Kadri, Wajid B.
Pakistan National Scientific & Technical Documentation Centre
Nazimabad, Karachi, Pakistan

Kaplan, Sidney
Astro-Electronic Products Div.
Radio Corp. of America
Princeton, N.J.

Kay, Lilly Lesin
2900 Adams Mill Rd., N.W.
Washington 9, D.C.

Kaye, Myra
Israel Atomic Energy Lab.
P.O. Box F056,
Tel-Aviv, Israel

Kee, Walter A.
The Martin Co.
Middle River, Mail J398
Baltimore 3, Md.

Keegan, Patricia
Smith, Kline & French Labs.
1500 Spring Garden St.
Philadelphia 1, Pa.

Keenan, Stella
Zinc Development Association
34 Berkeley Square
London W.1, England

Keim, C.P.
Technical Information Div.
Oak Ridge National Lab.
P.O. Box X
Oak Ridge, Tenn.

Kelly, Elizabeth C.
Bureau of Medicine
Food and Drug Administration
Department of Health, Education, and Welfare
Washington 25, D.C.

Kelner, R.C.
Computer Products Div.
Lab. for Electronics, Inc.
141 Maiden St.
Boston 18, Mass.

Kennedy, Robert Alvin
Technical Information Libraries
Murray Hill Lab.
Bell Telephone Labs., Inc.
Murray Hill, N.J.

Kent, Allen
Center for Documentation and Communication Research
School of Library Science
Western Reserve University
Cleveland 6, Ohio

Kenton, Charlotte
National Institutes of Health Library
Bethesda 14, Md.

Kessel, Benjamin
Computer Control Co., Inc.
92 Broad St.
Wellesley 57, Mass.

Kessler, M.M.
Lincoln Lab.
Massachusetts Institute of Technology
P.O. Box 73
Lexington 73, Mass.

Kessler, S.Sim
Scientific Literature Consultants
20 South 15th St.
Philadelphia 2, Pa.

Killian, Thomas J.
Office of Naval Research
Department of the Navy
Washington 25, D.C.

Kimura, Kazuo K.
Clinical Research Div.
U.S. Army Chemical Warfare Labs.
Army Chemical Center, Md.

King, Alexander
European Productivity Agency
3 rue Andre-Pascal
Paris 16, France

King, Gilbert W.
I.B.M. Research Center
Box 218
Yorktown Heights, N.Y.

Kingery, Robert E.
The New York Public Library, Att. CPD
Grand Central Station
P.O. Box 1932
New York 17, N.Y.

Kingston, John T.
Documentation, Inc.
2521 Connecticut Ave., N.W.
Washington 8, D.C.

Kinney, Margaret M.
Veterans Administration Hospital
Library
130 West Kingsbridge Rd.
Bronx 68, N.Y.

Kirsch, Russell A.
Applications Engineering Section
Data Processing Systems Div.
National Bureau of Standards
Washington 25, D.C.

Klever, Eugen
Chemisches Zentralblatt
Berlin-Schoeneberg, Hewaldstr. 5
Berlin, U.S.A. Sector, Germany

Knox, Margaret Enid
Dept. of Reference and Bibliography
University of Florida Libraries
Gainesville, Fla.

Knox, William T.
Technical Information Div.
Esso Research & Engineering Co.
P.O. Box 51
Linden, N.J.

Kobayashi, Yutaka
Investigation Section
The Japan Information Center of Science and Technology
Tokyo, Japan

Koehneke, J.H.
The Upjohn Co., Dept. 716
301 Henrietta St.
Kalamazoo, Mich.

Koelewijn, G.J.
Netherlands Patent Office
6 Willem Witsenplein
The Hague, Netherlands

Koenig, Elizabeth
Reference Library
National Institutes of Health
Bethesda 14, Md.

Kohler, Lois Esse
Library
U.S. Naval Engineering Experiment Station
Annapolis, Md.

Kolb, Margaret C.
Research Records
Merck Sharp & Dohme Research Labs.
Rahway, N.J.

Koller, Herbert R.
U.S. Patent Office
Department of Commerce
Washington 25, D.C.

Koonz, J.C.
Industrial Div.
The Magnavox Co.
Fort Wayne 4, Ind.

Kooser, Edwin deT.
Somerset County Guidance Center
217 West Main St.
Somerville, N.J.

Kortendick, James J.
Dept. of Library Science
The Catholic University of America
Washington 17, D.C.

Kotani, Masao
National Committee for Documentation
Science Council of Japan
Ueno Park, Tokyo, Japan

Kraft, Rose
Nortronics, Div. of Northrop Aircraft, Inc.
Hawthorne, Calif.

Krasovee, Frank
Smith, Kline & French Labs.
1500 Spring Garden St.
Philadelphia 1, Pa.

Kreithen, Alexander
Documentation, Inc.
2521 Connecticut Ave., N.W.
Washington 8, D.C.

Krekeler, George B.
Huntsville Computer Center
General Electric Co.
Box 988
Huntsville, Ala.

Krotz, Donald P.
California Research Corp.
200 Bush St.
San Francisco 4, Calif.

Krupp, Robert G.
Research Div. Library
American Cyanamid Co.
Bound Brook, N.J.

Kuentzel, Lester E.
Wyandotte Chemicals Corp.
Wyandotte, Mich.

Kuipers, John W.
Itek Corp.
1605 Trapelo Rd.
Waltham 54, Mass.

Kujawski, Wilma F.
Distillation Products Industries
Div. of Eastman Kodak Co.
755 Ridge Rd. West
Rochester 3, N.Y.

Kyle, Barbara
Social Sciences Documentation
7 Albemarle St.
London W. 1, England

Kyle, Robert J.
Technical Information Section
Georgia Institute of Technology
Atlanta 13, Ga.

L

Ladd, Robert Boyd
Operations Research Office
The Johns Hopkins University
6935 Arlington Rd.
Bethesda, Md.

Lage, Louise C.
The Lilly Research Labs.
Eli Lilly and Co.
740 South Alabama St.
Indianapolis 6, Ind.

La Hood, Charles G.
Photoduplication Service
The Library of Congress
Washington 25, D.C.

Lallemand, Jacqueline G.
Bibliothèque Royal de Belgium
5 rue du Musée
Brussels, Belgium

Lalli, Thomas J.
U.S. Army Signal Research & Development Lab.
Fort Monmouth, N.J.

Lamb, Sydney MacDonald
Dept. of Linguistics
University of California
Berkeley 4, Calif.

Lambert, J.M.
Glass Research Labs.
Pittsburgh Plate Glass Co.
Box 11472
Pittsburgh 38, Pa.

La Mer, Victor K.
Chandler Bldg.
Columbia University
New York 27, N.Y.

Lane, Wyvona A.
Materials Advisory Board
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

Lanham, Benjamin E.
Office of Research and Development
U.S. Patent Office
Department of Commerce
Washington 25, D.C.

Larkey, Sanford V.
Welch Medical Library
The Johns Hopkins University
1900 East Monument St.
Baltimore 5, Md.

Larrimer, Walter H.
Handbook of Biological Data
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

Lasareff, V.A.
Centre d'Études Nucleaires de Saclay
31 rue Belliard
Brussels, Belgium

Lawson, Louis R., Jr.
Charleston Research Lab.
West Virginia Pulp & Paper Co.
Charleston, S.C.

Lazarow, Arnold
Dept. of Anatomy
University of Minnesota Medical School
Minneapolis 14, Minn.

Lazarus, Rosalind
Arany Lorand Memorial Library
Hillside Hospital
P.O. Box 38
Glen Oaks, N.Y.

Lazerow, Samuel
National Library of Medicine
7th St. & Independence Ave., S.W.
Washington 25, D.C.

Leach, Marion
National Research Council of Canada Library
100 Sussex St.
Ottawa 2, Ontario, Canada

Leake, Chauncey D.
Dept. of Pharmacology
The Ohio State University
Columbus 10, Ohio

Lear, George M.
Eastman Kodak Co.
1015 Wisconsin Ave., N.W.
Washington, D.C.

Lear, John
The Saturday Review
25 West 45th St.
New York 36, N.Y.

LeCompte, Stuart B., Jr.
New Jersey Agricultural Experiment Station
Rutgers University
New Brunswick, N.J.

Ledley, Robert S.
Engineering Dept.
407 Tompkins Hall
George Washington University
Washington, D.C.

Lee, Milton O.
Federation of American Societies for Experimental Biology
9650 Wisconsin Ave.
Washington 14, D.C.

Leibowitz, J.
U.S. Patent Office
Department of Commerce
Washington 25, D.C.

Leitch, Isabella
Commonwealth Bureau of Animal Nutrition
Rowett Research Institute
Bucksburn, Aberdeenshire, Scotland

LeRoy, Andre Julien
Centre d'Études Nucleaires de Saclay
Boite Postale 2, Gif-sur-Yvette
Seine et Oise, France

Levenson, Leonard L.
Mechanics Bldg.
University of California
Berkeley 4, Calif.

Libby, Willard F.
U.S. Atomic Energy Commission
Washington 25, D.C.

Liebesny, Felix
The British Aluminium Co., Ltd.
Norfolk House, St. James's Square
London S.W.1, England

Liebhold, Klaus G.
Technical Military Planning Operation Library
General Electric Co.
735 State St.
Santa Barbara, Calif.

Liepina, Rita G.
Office of Scientific Information
National Science Foundation
1951 Constitution Ave., N.W.
Washington 25, D.C.

Liley, P.E.
Thermophysical Properties Research Center
Purdue University
Lafayette, Ind.

Lindenmeyer, Harold Fred
Scientific Library
U.S. Patent Office
Department of Commerce
Washington 25, D.C.

Ling, Chih Bing
Academia Sinica
P.O. Box 63
Taichung, Taiwan (Formosa), China

Lipetz, Ben-Ami
Battelle Memorial Institute
505 King Ave.
Columbus 1, Ohio

Little, Gretchen D.
Atlas Powder Co. Library
Wilmington 99, Del.

Livingston, George A.
3001 North Stuart St.
Arlington 7, Va.

Lloyd, Jeanne Hereford
Citrus Experiment Station Library
University of California
Riverside, Calif.

Lloyd, Robert
Recordak Corp.
1 Wanamaker Place
New York, N.Y.

Locke, William N.
Massachusetts Institute of Technology Libraries
77 Massachusetts Ave.
Cambridge 39, Mass.

Loftus, Martin L.
Joint Bank Fund Library
International Monetary Fund
Washington 25, D.C.

Logsdon, Richard H.
Columbia University Libraries
New York 27, N.Y.

Long, Frank R.
Technical Information Dept.
Atomics International
P.O. Box 309
Canoga Park, Calif.

Longare, Dorothy
Medical Library
American Cancer Society, Inc.
521 West 57th St.
New York 19, N.Y.

Loomis, Robert
Data Processing Div.
Mathematics & Applications Dept.
I.B.M. Corp.
112 East Post Rd.
White Plains, N.Y.

Loosjes, Theodoor P.
Library of the Agricultural University
Wageningen, Netherlands

Lorback, R.D.
Eastman Kodak Co.
343 State St.
Rochester 3, N.Y.

Lotz, J.W.
E.I. du Pont de Nemours & Co., Inc.
Jackson Lab., Organic Chemicals Dept.
P.O. Box 525
Wilmington 99, Del.

Low, Ward C.
Advanced Research Projects Agency
Room 3D 170, The Pentagon
Washington, D.C.

Lowry, W.Kenneth
Murray Hill Lab.
Bell Telephone Labs., Inc.
Murray Hill, N.J.

Lucas, Edward J.
Crosley Div., Avoc Mfg. Co.
750 Commonwealth Av.
Boston, Mass.

Luhn, Hans Peter
I.B.M. Research Center
P.O. Box 218
Yorktown Heights, N.Y.

Lykoudis, Paul
School of Aeronautical Engineering
Purdue University
Lafayette, Ind.

Lyman, Dwight C.
U.S. Navy Underwater Sound Lab. Library
Fort Trumbull
New London, Conn.

Lyon, Cathryn C.
Technical Information Section
Applied Science Corp. of Princeton
P.O. Box 44
Princeton, N.J.

Lyon, Ethel V.
Portland Cement Assoc. Library
33 West Grand Ave.
Chicago 10, Ill.

M

MacCorquodale, D.W.
Scientific Div.
Abbott Labs.
North Chicago, Ill.

Machelson, Sonya G.
The Library of Congress
Washington 25, D.C.

Machovec, Charles R.
Los Alamos Scientific Lab. Library
1981-A North Rd.
Los Alamos, N. Mex.

Mack, James D.
Lehigh University Library
Bethlehem, Pa.

MacKay, Donald M.
Dept. of Physics
King's College
University of London
Strand, W.C. 2, London, England

MacKenzie, Robert A., Jr.
Operations, Wing B, T & I Bldg.
General Electric Computer Dept.
Arizona State College
Tempe, Ariz.

MacLaughlin, W.H.
Abbott Labs.
North Chicago, Ill.

MacLennan, John Scott
Research Div.
American Cyanamid Co.
Bound Brook, N.J.

MacMillan, Judith T.
Cardiovascular Literature Project
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

MacRae, Donald Fraser
Industrial Research Services Dept.
Ontario Research Foundation
43 Queen's Park Crescent
Toronto 5, Ontario, Canada

Magistrate, Laura A.
Imperial Chemical Industries (N.Y.) Ltd.
488 Madison Ave.
New York 22, N.Y.

Maizell, Robert E.
American Institute of Physics
335 East 45th St.
New York 17, N.Y.

Majewski, Z.
Central Institute of Scientific and Technical Documentation
A l, Niepodległości 188
Warsaw, Poland

Maloney, Clifford J.
Chemical Corps Research and Development Command
U.S. Army Biological Warfare Labs.
Fort Detrick, Frederick, Md.

Mandelbrot, Benoit
Faculté des Sciences de Lille
13 Place Philippe Lebon
Lille (Nord), France

Manning, Bayless
Yale University Law School
New Haven, Conn.

Mannini, Walter
Socony Mobil Oil Co., Inc.
Refinery Engineering Div. Library
150 East 42nd St.
New York 17, N.Y.

Marden, Ethel Cox
Applications Engineering Section
Data Processing Systems Div.
National Bureau of Standards
Washington 25, D.C.

Maron, Melvin E.
The Ramo-Wooldridge Corp.
5500 West El Segundo Blvd.
Los Angeles 45, Calif.

Marquand, Carl B.
U.S. Army Chemical Corps Advisory Council
Army Chemical Center, Md.

Marr, Eleanor B.
Hunter College
606 West 120th St.
New York 27, N.Y.

Martin, D.C.
The Royal Society
Burlington House, Piccadilly
London W. 1, England

Martin, Genevieve H.
Allen-Bradley Co., Technical Library
136 West Greenfield Ave.
Milwaukee 4, Wis.

Martin, LeRoy B., Jr.
The Service Bureau Corp.
50 Franklin St.
New York 22, N.Y.

Martin, Louis H.
Crosley Div.
Avco Mfg. Corp.
750 Commonwealth Ave.
Boston, Mass.

Martin, Robert L.
Quartermaster Research & Engineering Command Technical Library
Natick, Mass.

Masterman, Margaret
Cambridge Language Research Unit
20 Millington Rd.
Cambridge, England

Matejka, L.
The Computation Lab. of Harvard University
Cambridge 38, Mass.

Mathys, Nel
Library Services Branch
Rome Air Development Center
Griffiss Air Force Base
Rome, N.Y.

Maung, U.Zin
Technical Information Center
Union of Burma Applied Research Institute
1 Kanbe Rd., Yankin P.O.
Rangoon, India

Max, George D.
Technical Information and Intelligence Office
Headquarters, Air Force Cambridge Research Center
Laurence G.Hanscom Field, Bedford, Mass.

Maynard, Paul J.
Air Research Dept.
The Library of Congress
Washington 25, D.C.

Mayse, Richard D.
Continental Oil Co. Library
Drawer 1267
Ponca City, Okla.

McAllister, David T.
Technical Information Dept.
U.S. Naval Ordnance Test Station
China Lake, Calif.

McCall, Bryant
Apparatus and Optical Div.
Eastman Kodak Co.
400 Plymouth Ave., North
Rochester 4, N.Y.

McCann, Eleanor
Duquesne University Library
Pittsburgh 19, Pa.

McClenahan, William S.
Chemical and Physical Research Div.
The Standard Oil Co. (Ohio)
4440 Warrensville Center Rd.
Cleveland 28, Ohio

McCormick, E.M.
Naval Ordnance Lab.
Corona, Calif.

McGann, Paul W.
Bureau of Mines
U.S. Department of the Interior
Washington 25, D.C.

McGeever, Emmett B.
Science Library, Dabney Hall
University of Tennessee
Knoxville, Tenn.

McGoldrick, Patrick J.
Computer Dept.
General Electric Co.
1103 North Central Ave.
Phoenix, Ariz.

McGowan, John P.
Technological Institute Library
Northwestern University
Evanston, Ill.

McGuire, Robert E.
Office of Director of Research
Boeing Airplane Co.
P.O. Box 3707
Seattle 24, Wash.

McKenna, F.E.
Research and Engineering Dept.
Air Reduction Co., Inc.
Murray Hill, N.J.

McKenzie, Lawson M.
Naval Research Advisory Committee
Office of Naval Research
Washington 25, D.C.

McKlveen, John Robert
Chemical Div.
Research and Development Library
Celanese Corp. of America
P.O. Box 8
Clarkwood, Tex.

Medwied, M.
Israel Mining Industries Labs.
P.O. Box 5249
Haifa, Israel

Mehaffey, Jean S.
The Sherwin-Williams Co.
1082 East 141st St.
Cleveland 10, Ohio

Mellon, M.G.
Dept. of Chemistry
Purdue University
Lafayette, Ind.

Melville, Richard W.
Stanford Research Institute
Menlo Park, Calif.

Menzel, Herbert
Bureau of Applied Social Research
Columbia University
605 West 115th St.
New York 25, N.Y.

Meredith, G.Patrick
Dept. of Psychology
The University
Leeds 2, England

Messinger, Harley B.
University of California Medical School
San Francisco, Calif.

Mews, Hazel
Dept. of Librarianship
University of the Witwatersrand
Milner Park, Johannesburg, South Africa

Meyer, Jean Marc
Ciba Ltd.
Klybeckstr. 141
Basle, Switzerland

Meyers, Willard L.
Hawkeye Works
Eastman Kodak Co.
343 State St.
Rochester, N.Y.

Meyriat, Jean
International Committee for Social Sciences Documentation
27 rue Saint-Guillaume
Paris 7, France

Mikhailov, Aleksandr I.
All-Union Institute of Scientific and Technical Information
Academy of Sciences of the U.S.S.R.
Baltiiskii Poselok 42-B
Moscow 57, Russia

Mikhailov, Oleg.A.
All-Union Institute of Scientific and Technical Information
Academy of Sciences of the U.S.S.R.
Baltiiskii Poselok 42-B
Moscow 57, Russia

Miles, Walter Richard
Naval Medical Research Lab.
U.S. Naval Submarine Base
New London, Conn.

Millar, Elaine
Pergamon Press, Inc.
122 East 55th St.
New York 22, N.Y.

Miller, Clem O.
Div. of Chemistry and Chemical Technology
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

Miller, Eugene
Documentation, Inc.
2521 Connecticut Ave., N.W.
Washington 8, D.C.

Miller, Leonard E.
Technical Information Div.
California Research Corp.
576 Standard Ave.
Richmond, Calif.

Miller, Michael
Air Force Assoc.
Mills Bldg.
Washington 6, D.C.

Miller, Robert B.
Technical Planning Committee
I.B.M. Corp.
P.O. Box 390
Poughkeepsie, N.Y.

Miller, Robert D.
Technical Information Dept.
Shell Development Co.
3737 Bellaire Blvd.
Houston 25, Tex.

Mills, Daniel A.
General Foods Corp.
555 South Broadway
Tarrytown, N.Y.

Milmore, Oswald H.
Shell Development Co.
4560 Horton St.
Emeryville 8, Calif.

Minder, Thomas L.
Research Div.
Curtiss-Wright Corp.
Quehanna, Pa.

Minsky, Marvin L.
Dept. of Mathematics
Massachusetts Institute of Technology
Cambridge 39, Mass.

Miwa, Daisaku
Research and Information Bureau
The Science and Technics Agency
2-2 Kasumigaseki Chiyoda-ku
Tokyo, Japan

Moats, John S.
Armed Services Technical Information Agency
Arlington Hall Station
Arlington 12, Va.

Mocium, A.T.
Corporate Engineering Standards
North American Aviation, Inc.
International Airport
Los Angeles 45, Calif.

Mohrhardt, Foster E.
U.S. Department of Agriculture Library
Washington 25, D.C.

Molo, William L.
Det. 3, HQ., Air Weather Service
FOB-4, Suitland, Md.

Montgomery, Edward B.
Industrial Computer Center
General Electric Co.
Tempe, Ariz.

Mooers, Calvin N.
Zator Co.
140½ Mt. Auburn St.
Cambridge 38, Mass.

Moore, Alexander M.
Parke, Davis & Co., Inc.
Joseph Campau at the River
Detroit 32, Mich.

Moore, Edward F.
Murray Hill Lab.
Bell Telephone Labs., Inc.
Murray Hill, N.J.

Moore, Edythe
Technical Information Center
American Potash & Chemical Corp.
201 West Washington Blvd.
Whittier, Calif.

Moore, Ernest George
ARS Information Div.
U.S. Department of Agriculture
Washington 25, D.C.

Morf, Rudolph
International Union of Pure and Applied Chemistry
c/o Sandoz Ltd.
Basle 13, Switzerland

Morris, Albert V.
2320 20th St., N.W.
Washington, D.C.

Morris, Marlis M.
Research and Development Labs.
Foote Mineral Co.
Berwyn, Pa.

Morsch, Lucile
The Library of Congress
Washington 25, D.C.

Morse, Kenneth T.
New York Society of Electron Microscopists
2 East 63rd St.
New York 21, N.Y.

Morse, Philip M.
Dept. of Physics
Massachusetts Institute of Technology
Cambridge 39, Mass.

Mote, Lionel J.B.
"Shell" Research Ltd.
Thornton Research Centre
P.O. Box 1, Chester, Wirral
Cheshire, England

Mount, Ellis
Federal Telecommunication Labs. Library
500 Washington Ave.
Nutley 10, N.J.

Mulcahy, Bertram A.
National Advisory Committee for Aeronautics
Washington 25, D.C.

Murphy, James S.
The Rockefeller Institute Library Committee
66th St. and York Ave.
New York 21, N.Y.

Murphy, Robert W.
Special Engineering Product Div., I.B.M. Corp.
182 North Hamilton St.
Poughkeepsie, N.Y.

Murra, Kathrine O.
The Library of Congress
Washington 25, D.C.

Murray, Virginia E.
McGill University Library School
3459 McTavish St.
Montreal 2, Quebec, Canada

N

Nacke, Otto
Federation Internationale de Documentation
Staphenhorstr. 62
Bielefeld, West Germany

Nadaski, Miklos
Université de Montreal
Case Postale 6128
Montreal, Quebec, Canada

Nally, Maurice E.
Technical Publications
Research and Advanced Development Div.
Avco Mfg. Co.
20 South Union St.
Lawrence, Mass.

Natrella, Joseph V.
Computer Dept.
General Electric Co.
3240 Wilson Blvd.
Arlington, Va.

Naughton, F.
Executive Office Bldg.
The White House
Washington, D.C.

Needham, Roger M.
Cambridge Language Research Unit
20 Millington Rd.
Cambridge, England

Neeland, Frances
Remington Rand Univac
Sperry Rand Corp.
1900 West Allegheny Ave.
Philadelphia 29, Pa.

Neumann, Albrecht J.
Office of Naval Research
Department of the Navy
Washington 25, D.C.

Nevenzel, Judd C.
Atomic Energy Project
University of California
P.O. Box 24164
Los Angeles 24, Calif.

Newman, Simon M.
Office of Research and Development
U.S. Patent Office
Department of Commerce
Washington 25, D.C.

Nicholson, Natalie N.
Charles Hayden Memorial Library
Massachusetts Institute of Technology
Cambridge 39, Mass.

Nicolaus, John J.
Bureau of Ships Technical Library
Department of the Navy
Washington 25, D.C.

Nijhoff, H.F.
Office of the Scientific Attaché
Netherlands Embassy
1470 Euclid St., N.W.
Washington 9, D.C.

Nikitin, Peter
International Atomic Energy Agency
Vienna 1, Kaerntnerring, Austria

Niuison, Wade S.
Recordak Corp.
1 Wanamaker Place
New York, N.Y.

Nolan, John
Product Planning
I.B.M. Corp.
Monterey and Cottle Rds.
San Jose, Calif.

North, C.T.
Dept. of Professional Education
The National Foundation
800 Second St.
New York 17, N.Y.

North, Jeanne B.
United Aircraft Corp. Library
400 Main St.
East Hartford, Conn.

Nutting, Howard S.
Central Research Index
The Dow Chemical Co.
Midland, Mich.

O

Oakes, Ann
5936 Brook Drive
Falls Church, Va.

Oatfield, Harold
Chas. Pfizer & Co., Inc.
11 Bartlett St.
Brooklyn 6, N.Y.

O'Connor, John J.
Remington Rand Univac
Sperry Rand Corp.
1900 West Allegheny Ave.
Philadelphia 29, Pa.

Odbert, Henry S.
Biological and Medical Sciences
National Science Foundation
1951 Constitution Ave., N.W.
Washington 25, D.C.

Odishaw, Hugh
U.S. National Committee for IGY
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

Odorkiewicz, Barbara J.
Stamford Research Labs.
American Cyanamid Co.
Stamford, Conn.

Oettinger, Anthony G.
The Computation Lab. of Harvard University
Cambridge 38, Mass.

Ohlman, Herbert M.
System Development Corp.
2500 Colorado Ave.
Santa Monica, Calif.

Öhman, Einar
The Swedish Iron Masters' Assoc.
Jernkontoret Kungsträdgårdsgatan 6
Stockholm C, Sweden

O'Leary, Francis Bernard
Institute of Technology
University of Minnesota
Minneapolis 14, Minn.

Olejar, Paul D.
Technical Information Div.
U.S. Army Chemical Warfare Labs.
Army Chemical Center, Md.

Oliver, Margaret A.
Research Labs.
Rohm & Haas Co.
5000 Richmond St.
Philadelphia 37, Pa.

Olsen, James L., Jr.
Smith, Kline & French Labs.
1500 Spring Garden St.
Philadelphia 1, Pa.

Olsoni, Karl Emerik
5434 Broad Branch Rd., N.W.
Washington 15, D.C.

Oman, Paul Wilson
ARS, Entomology Research Div.
Plant Industry Station
U.S. Department of Agriculture
Beltsville, Md.

Opler, Ascher
Computer Usage Co.
18 East 41st St.
New York 17, N.Y.

Orchard, Earl
Special Engineering Product Div.
I.B.M. Corp.
182 North Hamilton St.
Poughkeepsie, N.Y.

Ordway, G.L.
Operations Research, Inc.
8605 Cameron St.
Silver Spring, Md.

Orlans, Harold
Office of Special Studies
National Science Foundation
1951 Constitution Ave., N.W.
Washington 25, D.C.

Ornsen, Kate C.
Research and Development Library
Sun Oil Co.
Box 426
Marcus Hook, Pa.

O'Rourke, D.T.
746 East 2nd St.
Brooklyn 18, N.Y.

Orr, Richard Howard
Institute for Advancement of Medical Communication
37 East 67th St.
New York 21, N.Y.

Orr, R.J.
Technical Information Section
Polymer Corp. Ltd.
Sarnia, Ontario, Canada

Ostvold, Harold O.
The New York Public Library
Grand Central Station, Att. S & T, P.O. Box 1932
New York 17, N.Y.

Otterman, Joseph
Willow Run Labs.
University of Michigan
Ann Arbor, Mich.

P

Pagès, Robert
Laboratoire Psychologie Sociale de la Sorbonne
2 rue du Plâtre
Paris 4, France

Pardes, Norman
Raytheon Mfg. Co.
Andover, Mass.

Paretti, Linda
National Committee for Production
Viale Regina Margherita 83/D
Rome, Italy

Parker, Dorothy
The Rockefeller Foundation
49 West 49th St.
New York 20, N.Y.

Parman, Lee F.
Library Div. 7221
Sandia Corp.
P.O. Box 5800
Albuquerque, N. Mex.

Patterson, William C.
Agricultural Research Service
Animal Disease and Parasite Research Div.
U.S. Department of Agriculture
Beltsville, Md.

Paulo, Zeferino Ferreira
Overseas Scientific Documentation Center
Rua da Yungueira 86
Lisbon, Portugal

Paulus, W.K.
Sandia Corp.
Sandia Base
Albuquerque, N. Mex.

Pavlov, V.A.
Embassy of the U.S.S.R.
1125 16th St., N.W.
Washington, D.C.

Peakes, Gilbert L.
276 Grove Ave.
Metuchen, N.J.

Peitsch, Elinor
Smith, Kline & French Labs.
1500 Spring Garden St.
Philadelphia 1, Pa.

Pentz, Clyde A., Jr.
Union Carbide Chemicals Co.
Development Dept.
South Charleston 3, W.Va.

Perez, Abraham A.
Tactical Systems Lab.
Litton Industries
1548 Granville Ave.
West Los Angeles 25, Calif.

Perry, James W.
Center for Documentation and
Communication Research
School of Library Science
Western Reserve University
Cleveland 6, Ohio

Pfeffer, Harold
U.S. Patent Office
Department of Commerce
Washington 25, D.C.

Phaneuf, Charles A.
Crosley Div.
Avco Mfg. Co.
Boston, Mass.

Phelps, Ralph H.
Engineering Societies Library
United Engineering Trustees, Inc.
29 West 39th St.
New York 18, N.Y.

Phillips, Van B.
Recordak Corp.
415 Madison Ave.
New York 17, N.Y.

Pietsch, Eric
Gmelin Institute
f. Anorganische Chemie u. Grenzglbrette
Varrentrappstr. 40/42
Frankfurt/Main, Germany

Pirlot, Paul Leon
Dept. of Biology
University of Montreal
Case Postale 6128
Montreal, Quebec, Canada

Piternick, George
Catalog Dept.
University of California
Berkeley 4, Calif.

Porter, J.R.
Journal of Bacteriology
State University of Iowa Medical Labs.
Iowa City, Iowa

Power, Eugene B.
University Microfilms
313 North 1st St.
Ann Arbor, Mich.

Pratt, Malcolm
108 Spring Drive
Falls Church, Va.

Pratt, V.Lorraine
Stanford Research Institute Library
Menlo Park, Calif.

Prell, Donald B.
Applications Engineering Dept.
Benson-Lehner Corp.
11930 West Olympic Blvd.
Los Angeles 64, Calif.

Prescott, G.W.
Dept. of Botany
Michigan State University
East Lansing, Mich.

Price, Derek J.
Institute for Advanced Study
Princeton, N.J.

Prochovnick, Ammiel
Research Information Service
The John Crerar Library
86 East Randolph St.
Chicago 1, Ill.

Protic, Z.D.
Yugoslovenski Centar za Technicku i Naucnu Documentaciyu
Admirala Geprata 16
Belgrade, Yugoslavia

Puleston, Dennis
Information Div.
Brookhaven National Lab.
Upton, Long Island, N.Y.

Putscher, Richard E.
Patent Div.
E.I. du Pont de Nemours & Co., Inc.
Wilmington 98, Del.

R

Rafferty, James A.
Broadview Research Corp.
1680 Wisconsin Ave., N.W.
Washington 7, D.C.

Rajchman, Jan A.
Radio Corp. of America Labs.
Princeton, N.J.

Randall, Gordon E.
Technical Information Branch
Aro, Inc.
Arnold Engineering Development Center
Tullahoma, Tenn.

Ranganathan, S.R.
University of Madras
775 Seventeenth Cross Rd.
Bangalore 3, India

Rankin, C.Alice
Nopco Chemical Co. Library
Harrison, N.J.

Rasmussen, S.H.
Research Files
International Bank for Reconstruction and Development
1818 H St., N.W.
Washington 25, D.C.

Rathbun, Edith N.
Public Health Service
Department of Health, Education, and Welfare
Washington 25, D.C.

Raymond, Detlev J.
Pergamon Institute
122 East 55th St.
New York 22, N.Y.

Redman, Helen F.
Los Alamos Scientific Lab.
P.O. Box 1663
Los Alamos, N. Mex.

Reingold, Nathan
National Archives and Records Service
Government Services Administration
8th and Pennsylvania Ave., N.W.
Washington 25, D.C.

Reinsmith, Gerald
Research and Development Div.
Ordnance Weapons Command
Rock Island, Ill.

Reynolds, Margaret N.
Canada Dept. of Agriculture Library
Science Service Bldg.
Carling Ave.
Ottawa, Ontario, Canada

Ribeiro, Fernanda Leite
Service de Informacões Técnico-Cientificas
Av. Gen. Justo, 171-4
Rio de Janeiro, Brazil

Rice, Chester T.
Computer Dept.
General Electric Co.
1103 North Central Ave.
Phoenix, Ariz.

Rice, Randall, G.
The Chemical Abstracts Service
The Ohio State University
Columbus 10, Ohio

Richmond, Phyllis A.
University of Rochester Library
Rochester 20, N.Y.

Richter, John Henry
Technical Documents Services
Willow Run Labs.
University of Michigan Research Institute
P.O. Box 2008
Ann Arbor, Mich.

Rigby, Malcolm
Meteorological Abstracts
P.O. Box 1736
Washington 13, D.C.

Rigby, Marion K.
Meteorological Abstracts
P.O. Box 1736
Washington 13, D.C.

Riker, Albert J.
Dept. of Plant Pathology
University of Wisconsin
Madison 6, Wis.

Risen, Stuart G.
United Nations Technical Assistance Administration
P.O. Box 20, Grand Central Station
New York 17, N.Y.

Ritchie, Marguerite
Nuclear Products-ERCO
Div. of ACF Industries, Inc.
P.O. Box 5
Riverdale, Md.

Robb, John H.
1841 Irving St., N.W.
Washington 10, D.C.

Roberts, Ammarette
Research and Development Library
Lone Star Gas Co.
301 South Harwood
Dallas 1, Tex.

Roberts, Frederick Browne
Engineering
36 Bedford St.
Strand, London W.C.2, England

Robinson, L.
Data Processing Div.
I.B.M. Corp.
112 East Post Rd.
White Plains, N.Y.

Robinson, Thelma P.
7908 Sleaford Place
Bethesda 14, Md.

Rockwell, Harriet E.
Smith, Kline & French Labs.
1530 Spring Garden St.
Philadelphia 1, Pa.

Rogala, Lita
Royal Swedish Academy of Engineering Sciences
8 East 69th St.
New York 21, N.Y.

Rogalski, Leonore
Universal Oil Products Co. Library
30 Algonquin Rd.
Des Plaines, Ill.

Rogers, David J.
New York Botanical Garden
Bronx Park
New York 58, N.Y.

Rogers, Frank B.
National Library of Medicine
7th St. & Independence Ave., S.W.
Washington 25, D.C.

Ronan, C.A.
The Royal Society
Burlington House, Piccadilly
London W.1, England

Rosmini, Cornelia M.
B.F.Goodrich Chemical Co. Library
3135 Euclid Ave.
Cleveland 15, Ohio

Rossell, Robert E.
Research & Engineering Council of the Graphic Arts Industry, Inc.
5728 Connecticut Ave., N.W.
Washington 15, D.C.

Rossini, Frederick D.
Carnegie Institute of Technology
Schenley Park
Pittsburgh 13, Pa.

Rubinstein, Eli A.
Training Branch
National Institute of Mental Health
Bethesda 14, Md.

Ruhf, H.Clayton
John H.Breck, Inc.
115 Dwight St.
Springfield 3, Mass.

Rumyantsev, Leonid Alexandrovich
Embassy of the U.S.S.R.
1125 16th St., N.W.
Washington, D.C.

Rupprecht, T.A.
Technical Documents Dept.
Missile Div.
Chrysler Corp.
P.O. Box 2628
Detroit 3, Mich.

Ryan, John W.
General Electric AGT Library System
Bldg. 305
Cincinnati 15, Ohio

S

Sabel, C.S.
United Kingdom Atomic Energy Authority
Harwell, Didcot
Berks, England

Sainsbury, Thomas R.
Technical Information Dept.
American Machine & Foundry Co.
200 Henry St., Box 889
Stamford, Conn.

St. George, Emery
Hydel, Inc.
99 First St.
Cambridge 41, Mass.

Sakal, Emile H.
Technical Information Dept.
Warner-Lambert Research Institute
Tabor Rd.
Morris Plains, N.J.

Samain, Jacques
Centre National de la Recherche Scientifique
74 rue des Saints-Pères
Paris, 7, France

Samson, Robert F.
Rome Air Development Center
Griffiss Air Force Base
Rome, N.Y.

Sanburn, Kathleen L.
New Zealand Scientific Liaison Office
1907 K St., N.W.
Washington, D.C.

Sandoval, Armando M.
Centre de Documentatión Científica y Técnica de México
Plaza de la Ciudadela 6
Mexico 1, D.F. Mexico

Sanford, Albert
1200 Pinecrest Circle
Silver Spring, Md.

Sauter, Hubert E.
Aircraft Nuclear Propulsion Dept.
Research Information Unit
General Electric Co.
P.O. Box 132
Cincinnati 15, Ohio

Savage, T.R.
I.B.M. Research Center
P.O. Box 218
Yorktown Heights, N.Y.

Savidge, David
Remington Rand Univac
Sperry Rand Corp.
1630 L St., N.W.
Washington, D.C.

Savitz, Gerald S.
Technical Information Dept.
Merck Sharp & Dohme Research Labs.
Rahway, N.J.

Schauer, J.D.
Sandia Corp.
Sandia Base
Albuquerque, N. Mex.

Schell, Emil D.
I.B.M. Corp.
90 Crestview Rd.
Mountain Lakes, N.J.

Scherberg, Max G.
Headquarters, Wright Air Development Center
Wright-Patterson Air Force Base, Ohio

Schiffman, Zelda
Management Analysis Branch
Div. of Business Operations
National Institutes of Health
Bethesda 14, Md.

Schlochauer, Marianne
Itek Corp.
1605 Trapelo Rd.
Waltham 54, Mass.

Schlundt, Esther M.
Purdue University
130 Waldron St.
West Lafayette, Ind.

Schmidt, Charles J.
Technical Information Group
Knowles Atomic Power Lab.
General Electric Co.
P.O. Box 1072
Schenectady, N.Y.

Schmitt, Samuel A.
I.B.M. Research Center
Poughkeepsie, N.Y.

Schneller, George H.
Wyeth Labs.
Div. of American Home Products Corp.
Box 8299
Philadelphia 1, Pa.

Schobinger, B.A.
Technical Information Div.
Esso Research & Engineering Co.
P.O. Box 51
Linden, N.J.

Schoeffel, Eugene W.
P.O. Box 68
Sterling Drug, Inc.
Rothschild, Wis.

Schoengold, Morris D.
Technical Information Div.
Esso Research & Engineering Co.
P.O. Box 51
Linden, N.J.

Schroedel, E.C.
Special Engineering Product Div.
I.B.M. Corp.
112 East Post Rd.
White Plains, N.Y.

Schück-Kolben, Gerhard P.
Research and Development Div.
Consolidation Coal Co.
Library, Pa.

Schultz, Claire K.
Remington Rand Univac
Sperry Rand Corp.
1900 West Allegheny Ave.
Philadelphia 29, Pa.

Schulze, Else L.
Technical Library
The Procter & Gamble Co.
M.A. & R. Bldg., Ivorydale
Cincinnati 17, Ohio

Scott, Peter R.
Microreproduction Lab.
Massachusetts Institute of Technology
Cambridge 39, Mass.

Scott, Raymond L.
The Norwich Pharmacal Co.
13-27 Eaton Ave.
Norwich, N.Y.

Scoville, Herbert, Jr.
Old Georgetown Pike
McLean, Va.

Sebeok, Thomas A.
Research Center in Anthropology, Folklore, and Linguistics
Rayl House, Indiana University
Bloomington, Ind.

Secondi, Juan Carlos
Oficina Sanitaria Panamericana
Ministerio de Asuntos Sociales y Salud Pública, Chaco
Universidad de Buenos Aires
Charcas 684, Piso 4
Buenos Aires, Argentina

See, Richard
Office of Scientific Information
National Science Foundation
1951 Constitution Ave., N.W.
Washington 25, D.C.

Seidell, Atherton
2301 Connecticut Ave., N.W.
Washington 8, D.C.

Selye, Hans
Institute of Experimental Medicine
University of Montreal
Montreal, Quebec, Canada

Severance, Robert W.
Air University Library
Maxwell Air Force Base
Montgomery, Ala.

Sewell, Winifred
Squibb Institute for Medical Research Library
E.R.Squibb & Sons Co.
5 Georges Rd.
New Brunswick, N.J.

Sexton, Anna M.
Div. of Labs, and Research
New York State Dept. of Health
New Scotland Ave.
Albany 1, N.Y.

Shaw, Ralph R.
Graduate School of Library Service
Rutgers University
New Brunswick, N.J.

Sheel, P.
Indian National Scientific Documentation Centre
National Physical Lab.
New Delhi 12, India

Shepard, Joseph W.
Minnesota Mining & Mfg. Co.
600 Bush Ave.
St. Paul 6, Minn.

Shepherd, Clayton A.
Institute for Urban Studies
University of Pennsylvania
Philadelphia 4, Pa.

Sheppard, Mary McC.
Internat. Conf. on Scientific Information
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

Shera, Jesse H.
School of Library Science
Western Reserve University
Cleveland 6, Ohio

Sherman, May
National Heart Institute
National Institutes of Health
Bethesda 14, Md.

Sherrod, John
Science and Technology Div.
The Library of Congress
Washington 25, D.C.

Shortley, George
Booz-Allen Applied Research, Inc.
4921 Auburn Ave.
Bethesda, Md.

Shuey, Richard Lyman
Information Studies
General Electric Research Lab.
P.O. Box 1088
Schenectady, N.Y.

Shuttleworth, R.G.
South African Scientific Liaison Office
1907 K St., N.W.
Washington 6, D.C.

Simon, W.R.
Technical Information, Epidemiology Div.
Department of National Health & Welfare
808 Jackson Bldg.
Ottawa, Ontario, Canada

Simpson, F.W.
Technical Information Services
General Atomic, Div. of
General Dynamics Corp.
P.O. Box 608
San Diego 12, Calif.

Sinai, Martha
Academic Press, Inc.
111 Fifth Ave.
New York 3, N.Y.

Singer, T.E.R.
52 Vanderbilt Ave.
New York 17, N.Y.

Sisco, Frank T.
Engineering Foundation
29 West 39th St.
New York 18, N.Y.

Sklansky, Jack
Radio Corp. of America Labs.
Princeton, N.J.

Skolnik, Herman
Technical Information Div.
Hercules Powder Co. Research Center
Wilmington 99, Del.

Slee, Virgil N.
Commission on Professional and
Hospital Activities, Inc.
First National Bldg.
Ann Arbor, Mich.

Sloane, Alice L.
Cumulative Catalogue
The Library of Congress
Washington 25, D.C.

Smith, Andrenette F.
Technical Reference Library
Naval Medical Research Institute
National Naval Medical Center
Bethesda 14, Md.

Smith, Charles G.
U.S. Patent Office
Department of Commerce
Washington 25, D.C.

Smith, Donald M.
Polychemicals Dept.
Experimental Station
E.I. du Pont de Nemours & Co., Inc.
Wilmington 98, Del.

Smith, G.Donald
State College of Washington Library
Pullman, Wash.

Smith, J.Roland
United Kingdom Atomic Energy Authority
Library, Research and Development Br.
Risley, Warrington
Lanes, England

Smith, Maurice H.
The James Forrestal Research Center Library
Princeton University
Princeton, N.J.

Smith, May K.
Medical Research Div.
Esso Research & Engineering Co.
P.O. Box 45
Linden, N.J.

Smith, Robert B.
Eastman Kodak Co. Research Labs.
Bldg. 59, Kodak Park
Rochester 4, N.Y.

Smookler, Idair
Union Carbide Chemicals Co. Research Dept.
P.O. Box 8004
South Charleston 3, W.Va.

Sohns, Ernest R.
6504 Brook's Hill Court
Washington 16, D.C.

Sophar, Gerald J.
The Microcard Corp.
1725 N St., N.W.
Washington 6, D.C.

Southern, W.A.
Abbott Labs.
North Chicago, Ill.

Spärck-Jones, Karen
Cambridge Language Research Unit
20 Millington Rd.
Cambridge, England

Spengler, Kenneth C.
American Meteorological Society
3 Joy St.
Boston 8, Mass.

Sperl, Virginia R.
Albert Einstein College of Medicine Library
Eastchester Rd. & Morris Park Ave.
New York 61, N.Y.

Spiegelthal, Edwin S.
Computer Dept.
General Electric Co.
7235 Wisconsin Ave., Suite 201
Bethesda, Md.

Spiegler, Donna F.
Operations Research Group
Office of Naval Research
Washington 25, D.C.

Spinning, Mira E.
Bristol Labs., Inc. Library
P.O. Box 657
Syracuse, N.Y.

Spirit, Jiří
State Research Institute for
Material and Technology
Opletalova 25, Prague 3, Czechoslovakia

Spitzer, Ernest F.
Chas. Pfizer & Co., Inc.
11 Bartlett St.
Brooklyn 6, N.Y.

Sprod, Thomas D.
Australian Consulate General
636 5th Ave.
New York 20, N.Y.

Sprow, Allen J.
University of Pennsylvania
3815 Walnut St.
Philadelphia 4, Pa.

Spurr, Stephen H.
School of Natural Resources
University of Michigan
Ann Arbor, Mich.

Stearns, John F.
Headquarters, Air Research and
Development Command
Andrews Air Force Base
Washington 25, D.C.

Steeves, Henry Alan
RIAS, Inc.
7212 Bellona Ave.
Baltimore 12, Md.

Stein, Dimitri R.
1074 Washington Ave.
Pelham Manor, N.Y.

Steinecke, William M.
Research Dept.
Bethlehem Steel Co.
Bethlehem, Pa.

Steinmetz, Nell S.
Pacific Aeronautical Library
7660 Beverly Blvd.
Los Angeles 36, Calif.

Stern, Joshua
Office of Basic Instrumentation
National Bureau of Standards
Washington 25, D.C.

Sternberg, Virginia A.
Bettis Atomic Power Div.
Westinghouse Electric Corp.
P.O. Box 1468
Pittsburgh 30, Pa.

Stevens, Charles H.
School of Mechanical Engineering
Purdue University
Lafayette, Ind.

Stevens, Mary Elizabeth
Applications Engineering Section
Data Processing Systems Div.
National Bureau of Standards
Washington 25, D.C.

Stevenson, Chris G.
Hanford Labs. Operation
General Electric Co.
Richland 1, Wash.

Stiles, H.Edmund
Brooklyn Bridge Rd.
Laurel, Md.

Still, Joseph W.
School of Medicine
George Washington University
1335 H St., N.W.
Washington, D.C.

Stimmel, William Dale
Research Labs. Library
Allis Chalmers Mfg. Co.
Milwaukee 1, Wis.

Stockdale, Noel
The Australian National University Library
Box 4, G.P.O.
Canberra A.C.T., Australia

Stoll, Charles B.
John Wiley & Sons, Inc.
440 4th Ave.
New York 16, N.Y.

Stoll, Norman R.
The Rockefeller Institute
66th St. and York Ave.
New York 21, N.Y.

Stonier, Thomas
The Rockefeller Institute
66th St. and York Ave.
New York 21, N.Y.

Stork, Louis C.
Technical Literature Research Dept.
The Texas Co.
P.O. Box 509
Beacon, N.Y.

Stotz, E.H.
Dept. of Biochemistry
Medical Center, University of Rochester
Rochester 20, N.Y.

Stover, Vera L.
Research Council of Alberta
87th Ave. & 114th St.
Edmonton, Alberta, Canada

Strain, Paula M.
Air Research Div.
The Library of Congress
Washington 25, D.C.

Strandh, Sigvard
Office of Scientific Attache
Swedish Embassy
2249 R St., N.W.
Washington, D.C.

Strauss, Lucille J.
Chemistry and Physics Library
Whitmore Lab.
The Pennsylvania State University
University Park, Pa.

Sullivan, Ralph H.
Special Management Research Service
General Electric Co.
Crotonville, Ossining, N.Y.

Sunderlin, C.E.
European Research Associates
95 rue Gatti de Gammond
Brussels, Belgium

Swanson, Don R.
Control Systems Div.
The Ramo-Woolridge Corp.
P.O. Box 45534, Airport Station
Los Angeles 45, Calif.

Swartley, William Moyer
Pennsylvania Junior Republic Assoc.
Grove City, Pa.

Symon, Maxwell Stanley
Electric Regulator Corp.
Pearl St.
Norwalk, Conn.

T

Taggart, Rebecca L.
Aeronautical Engineering Library
Purdue University
Lafayette, Ind.

Taine, Seymour I.
National Library of Medicine
7th St. & Independence Ave., S.W.
Washington 25, D.C.

Talkington, Lester
I.B.M. Corp.
590 Madison Ave.
New York 22, N.Y.

Tanimoto, Taffee
Data Processing Div.
I.B.M. Corp.
112 East Post Rd.
White Plains, N.Y.

Tasman, Paul
I.B.M. World Trade Corp.
821 United Nations Plaza
New York 17, N.Y.

Tate, Fred A.
Administrative Engineering Dept.
General Motors Corp.
Box 188, North End Station
Detroit 2, Mich.

Taube, Mortimer
Documentation, Inc.
2521 Connecticut Ave., N.W.
Washington 8, D.C.

Tauber, Maurice F.
School of Library Service
Columbia University
New York 27, N.Y.

Taylor, Jean K.
Business and Technology Dept.
Cleveland Public Library
325 Superior Ave.
Cleveland 14, Ohio

Taylor, Kanardy L.
The Library of the Transportation Center
Northwestern University
1810 Hinman Ave.
Evanston, Ill.

Taylor, Robert L.
Engineering Dept.
E.I. du Pont de Nemours & Co., Inc.
Wilmington 98, Del.

Taylor, Robert S.
Lehigh University Library
Bethlehem, Pa.

te Nuyl, Th.W.
N.V. de Bataafsche Petroleum Maatschappij
Carel van Bylandtlaan 30
The Hague, Netherlands

Theriault, Frederic R.
7301 16th Place
Hyattsville, Md.

Thomas, Clayton L.
Harvard University School of Public Health
Boston, Mass.

Thompson, Mendall S.
Science Div.
The Library of Congress
Washington 25, D.C.

Thorensen, Ragnar
The Magnavox Co. Research Labs.
2255 South Carmelina Ave.
Los Angeles 64, Calif.

Tolkan, Constance H.
Office of Scientific Information
National Science Foundation
1951 Constitution Ave., N.W.
Washington 25, D.C.

Tolman, Mason
The New York State Library
Albany 1, N.Y.

Tonini, Dino
Societa Àdriatica di Elettricità
Ufficío Studi Campo Fava 552 F.
Venice, Italy

Toth, Alexander B.
4969 Allan Rd., N.W.
Washington, D.C.

Touloukian, Y.S.
School of Mechanical Engineering
Purdue University
Lafayette, Ind.

Townsend, John R.
Office of Assistant Secretary of Defense
Research and Engineering
The Pentagon, Washington 25, D.C.

Trager, George L.
Dept. of Anthropology & Linguistics
University of Buffalo
3435 Main St.
Buffalo 14, N.Y.

Tukey John W.
Dept. of Mathematics
Fine Hall
Princeton University
Princeton, N.J.

Tung, Yueh-Chiung
Library School
Columbia University
New York 27, N.Y.

Turanski, W.J.
Applications Research Center
Remington Rand Univac
Sperry Rand Corp.
19th St. & West Allegheny Ave.
Philadelphia 29, Pa.

Turnbull, Florence W.
Engineering Library
Sperry Gyroscope Co.
Great Neck, N.Y.

Turner, E.S.
Central Research Dept.
Minnesota Mining & Mfg. Co.
2301 Hudson Rd.
St. Paul 9, Minn.

Turner, Florence E.
School of Medicine
Boston University
80 East Concord St.
Boston 18, Mass.

U

Ullman, Guy
RIAS, Inc.
7212 Bellona Ave.
Baltimore 12, Md.

Ullmann, Hans Charles
Armed Services Technical Information Agency
Arlington Hall Station
Arlington 12, Va.

Urquhart, D.J.
Dept. of Scientific and Industrial Research
Lending Library Unit
20 Chester Terrace, Regents Park
London N.W.1, England

V

Valeri, Virginia R.
Arthur D.Little, Inc.
15 Acorn Park
Cambridge 40, Mass.

Vanags, E.M.
Aluminium Labs., Ltd.
P.O. Box 250
Arvida, Quebec, Canada

Vance, John K.
10317 Dettrick Ave.
Kensington, Md.

van Waasbergen, Pieter
International Patents Institute
97 Nieuwe Parklaan
The Hague, Netherlands

van Woerkom, A.J.
Electric Boat Div.
General Dynamics Corp.
P.O. Box 82
Niantic, Conn.

Vasconcelos, Maria Laurinda Miltao
Fernandas
Centre de Documentação
Cientifica Ultramarina
Rua da Junqueira 86
Lisbon, Portugal

Vernon, Janet
Whirlpool Corp.
300 Broad St.
St. Joseph, Mich.

Vianna Dias, Mario Ulisses
Institute Brasileiro de Bibliografia
e Documentação
Avenida General Justo 171-4° Andar
Rio de Janeiro, D.F., Brazil

Vickery, B.C.
Imperial Chemical Industries, Ltd.
Akers Research Labs.
The Frythe
Welwyn, Herts, England

Viele, Charles E.
Charleston Research Lab.
West Virginia Pulp & Paper Co.
Charleston, S.C.

Visscher, Maurice B.
Dept. of Physiology
University of Minnesota
Minneapolis 14, Minn.

Von Sonn, V.B.
Technical Information Services
The Rand Corp.
1700 Main St.
Santa Monica, Calif.

W

Wachtel, Irma S.
Council for Economic & Industry Research, Inc.
734 15th St., N.W.
Washington 5, D.C.

Waddell, Mary F.
Communicable Disease Center Library
50 Seventh St., N.E.
Atlanta 23, Ga.

Waddington, Guy
Office of Critical Tables
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

Wakerling, R.K.
Information Div.
University of California Radiation Lab.
Berkeley 4, Calif.

Walcott, M.Alena
Brandeis University Library
Waltham 54, Mass.

Waldo, W.H.
Organic Chemistry Div.
Monsanto Chemical Co.
1700 South Second St.
St. Louis 4, Mo.

Walker, William F.
Eastman Kodak Co. Research Labs.
Rochester 4, N.Y.

Wall, Eugene
Engineering Dept.
E.I. du Pont de Nemours & Co., Inc.
1007 Market St.
Wilmington 98, Del.

Wall, James R.
Alcoa Research Labs.
Aluminum Co. of America
P.O. Box 497
East St. Louis, Ill.

Walter, John E.
524 West Great Falls St.
Falls Church, Va.

Ward, Dorothy M.
Texicological Information Center
Chemical Warfare Labs.
Army Chemical Center, Md.

Ward, Gordon B.
Office of Scientific Information
National Science Foundation
1951 Constitution Ave., N.W.
Washington 25, D.C.

Ward, R.R.
Technical Information Services
Shell Development Co.
Emeryville 8, Calif.

Warheit, I.Albert
Div. of Information Services
U.S. Atomic Energy Commission
Washington 25, D.C.

Warren, Peter A.
Excerpta Medica Foundation
New York Academy of Medicine Bldg.
2 East 103rd St.
New York 29, N.Y.

Waters, John J.
Rome Air Development Center
Griffiss Air Force Base
Rome, N.Y.

Way, Katherine
Nuclear Data Group
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

Weaver, C.R.
Ohio Agricultural Experiment Station
Wooster, Ohio

Webber, John Putnam
U.S. Weather Bureau Library
Washington 25, D.C.

Weber, David C.
Harvard University Library
Cambridge 38, Mass.

Weed, Katherine K.
Bureau of Ordnance Library
Department of the Navy
Washington 25, D.C.

Weeks, Elizabeth H.
Research Div. Library
Raytheon Mfg. Co.
Waltham 54, Mass.

Weeks, Gene M.
Science Press, Inc.
23 West 45th St.
New York, N.Y.

Well, Barbara H.
Information Services Section
Metals Research Labs.
Electro Metallurgical Co.
4625 Royal Ave.
Niagara Falls, N.Y.

Weil, B.H.
Technical Information Div.
Esso Research & Engineering Co.
P.O. Box 51
Linden, N.J.

Weiner, Jack
The Institute of Paper Chemistry
1043 East South River St.
Appleton, Wis.

Weinstein, Shirley Jane
Research Div., Patent Section
Armour and Co.
1425 West 42nd St.
Union Stockyards, Chicago 9, Ill.

Weisman, Herman M.
Colorado State University Research Foundation
Fort Collins, Colo.

Weitlauf, Frederica M.
Research Library
Inland Steel Co.
East Chicago, Ind.

Weitz, R.G.
Technical Information Branch
Air Force Ballistic Missile Div.
Headquarters, Air Research and
Development Command
P.O. Box 262
Inglewood, Calif.

Weller, O.G.
Uganda Development Corp.
Technical Development Div.
P.O. Box 18
Entebbe, Uganda
British East Africa

Welt, Isaac D.
Cardiovascular Literature Project
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

Wendelschafer, Mary
The Upjohn Co. Library
301 Henrietta St.
Kalamazoo, Mich.

Werenbaum, J.W.
General Electric Computer Lab.
951 Commercial St.
Palo Alto, Calif.

Wescott, Dorothy S.
Rohm & Haas Co. Library
5000 Richmond St.
Philadelphia 37, Pa.

Wessel, Carl J.
Prevention of Deterioration Center
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

Whaley, Fred R.
Linde Air Products Co.
Tonawanda, N.Y.

Whatmough, Joshua
Dept. of Linguistics
54 Dudley Hall
Harvard University
Cambridge 38, Mass.

White, Peregrine
U.S. Army Office of Ordnance Research
Box CM, Duke Station
Durham, N.C.

White, Sophia P.
Electrodata Div. Library
Burroughs Corp.
460 Sierra Madre Villa
Pasadena 8, Calif.

Whitlock, J.W.
Board of Trustees
Cambridge Language Research Unit
Cambridge, England

Whitman, Bradley
Schering Corp.
60 Orange St.
Bloomfield, N.J.

Whitnah, Margaret C.
The Ramo-Wooldridge Corp.
P.O. Box 90453, Airport Station
Los Angeles 45, Calif.

Wickline, Marian E.
The Dow Chemical Co.
P.O. Box 351
Pittsburgh, Calif.

Wildhack, W.A.
Office of Basic Instrumentation
National Bureau of Standards
Washington 25, D.C.

Wile, Annadel N.
2701 Connecticut Ave., N.W.
Washington 8, D.C.

Wilkinson, William Archer
Monsanto Chemical Co.
1700 South 2nd St.
St. Louis 4, Mo.

Williams, Josephine
Office of Information
National Academy of Sciences
2101 Constitution Ave., N.W.
Washington 25, D.C.

Williams, Thyllis
Itek Corp.
1605 Trapelo Rd.
Waltham 54, Mass.

Williamson, Harry
Canadian Scientific Liaison Office
1907 K St., N.W.
Washington 6, D.C.

Wilson, A.J.C.
Dept. of Physics
University College
Cardiff, Great Britain

Wilson, C.W.J.
United Kingdom Atomic Energy Authority
Harwell, Didcot
Berks, England

Wilson, John H., Jr.
Atlantic Research Corp.
Alexandria, Va.

Winn, Sidney E.
Technical Information Section
Gulf Research & Development Co.
P.O. Drawer 2038
Pittsburgh 30, Pa.

Wise, Leonard G.
Central Research Dept.
Experimental Station
E.I. du Pont de Nemours & Co., Inc.
Wilmington 98, Del.

Wixon, Darvey W.
U.S. Army Signal Research and Development Lab.
Fort Monmouth, N.J.

Woelfel, Helen L.
U.S. Army Medical Research Lab.
Fort Knox, Ky.

Wolfle, Dael
American Association for the Advancement of Science
1515 Massachusetts Ave., N.W.
Washington 5, D.C.

Woodbury, William W.
I.B.M. Corp. Research Labs.
Monterey and Cottle Rds.
San Jose, Calif.

Woodhull, E.H.
Haloid Xerox, Inc.
2 Haloid St.
Rochester 3, N.Y.

Wooster, Harold
Air Force Office of Scientific Research
Washington 25, D.C.

Worsley, Peter K.
Benson-Lehner Corp.
11930 West Olympic Blvd.
Los Angeles 64, Calif.

Wright, Meredith S.
National Carbon Research Labs. Library
P.O. Box 6116
Cleveland 1, Ohio

Wright, R.C.
Royal Aircraft Establishment Library
South Farnborough
Hants, England

Wyart, Jean
Université de Paris, Sorbonne
1 rue Victor-Cousin
Paris 5, France

Wyllie, John Cook
Alderman Library
University of Virginia
Charlottesville, Va.

Wylly, Alexander
Planning Research Corp.
10970 Le Conte Ave.
Los Angeles 24, Calif.

Y

Yngve, Victor H.
Research Lab. of Electronics
Massachusetts Institute of Technology
Cambridge 39, Mass.

Yochelson, Ellis L.
U.S. Geological Survey
U.S. National Museum
Washington 25, D.C.

Young, Hoylande D.
Argonne National Lab.
P.O. Box 299
Lemont, Ill.

Young, Robert W.
U.S. Navy Electronics Lab.
San Diego 52, Calif.

Yovits, Marshall C.
Office of Naval Research
Information Systems Br.
Department of the Navy
Washington 25, D.C.

Z

Zarbock, James W.
Oxford University Press
417 5th Ave.
New York 16, N.Y.

Zarember, I.
Technical Information Div.
Esso Research & Engineering Co.
P.O. Box 51
Linden, N.J.

Zhdanov, Yryi Andreevich
All-Union Institute of Scientific and Technical Information
Academy of Sciences of the U.S.S.R.
Baltiiskii Poselok 42-B
Moscow 57, U.S.S.R.

Zobrist, Benedict K.
Technical Information Branch 2250
Rock Island Arsenal
Rock Island, Ill.

Zweifel, LeRoy G.
College of Engineering
University of Wisconsin
Madison, Wis.

Zwemer, Raymund L.
Office of the Science Advisor
Department of State
Washington 25, D.C.

Zwolinski, Bruno J.
Chemical & Petroleum Research Lab.
Carnegie Institute of Technology
Pittsburgh 13, Pa.

INDEX

A NOTE ON THE TYPOGRAPHY

The graceful and legible type face in which this book is composed is Monotype Bembo. The roman was adapted in 1929 from a fifteenth century model first used by the publisher, Aldus Manutius, for a pamphlet by a young Renaissance scholar, Pietro Bembo. The italic used here was developed some years after the roman appeared. It is based on designs used in Venice after 1524 by the writing-master Tagliente. Bembo is a popular face for literary work, particularly abroad, but it has not often been used for scientific material.

The text of this book has been composed with the expectation that electronic computers will be used in analyzing the material for storage and search. Indeed, results of one such experiment were shown among the exhibits during the Conference. The skilled reader may have noted unusually wide spacing between sentences. This results from use of double spaces after sentence stops to enable electronic computers easily to discriminate sentence endings from periods used in figures, abbreviations, and so forth.

The punched tapes from the monotype keyboard machines from which type for this book was cast have been preserved for other computer experiments. Arrangements to borrow the tapes may be made through the National Academy of Sciences.