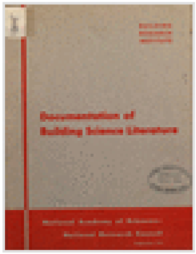

Documentation of Building Science Literature (1960)

Pages
54

Size
8.5 x 10

ISBN
0309361311

Building Research Institute; Division of Engineering and Industrial Research; National Research Council

Documentation of Building Science Literature

Proceedings of a program conducted
as part of the 1959 Fall Conferences
of the Building Research Institute,
Division of Engineering and Industrial Research

Publication 791
NATIONAL ACADEMY OF SCIENCES-NATIONAL RESEARCH COUNCIL
Washington, D. C.
1960

2-11

Inquiries regarding the the contents of this publication or the BRI 1959 Fall Conferences, of which this program was a part, may be directed to the Publications Department, Building Research Institute, Division of Engineering and Industrial Research, National Academy of Sciences-National Research Council.

Available from the
Printing and Publishing Office
National Academy of Sciences-National Research Council
Washington 25, D. C.

Price, $2.00

Library of Congress Catalog Card No.: 60-60057

CONTENTS

Need for Better Access to Building Science Literature

By Glenn H. Beyer,* Director,
Housing Research Center, Cornell University

The term "building science" is of sufficiently recent vintage that I would like to define it at the outset. "Building," as it will be used in this paper, encompasses all aspects of man's shelter, including the physical, social and economic. Obviously, this is a much broader definition than one which might limit its meaning to the isolated process of erecting a structure. "Science," as used here, reflects an encompassing of the physical, biological and social sciences as they have relationships to building. The definition of building science is not, therefore, limited to the narrow technological concept it was a few years ago.

Although this broader concept of building science has been accepted by many and was, in fact, the subject for a Building Research Advisory Board conference in 1953, it cannot be assumed that it is commonly accepted by all groups and individuals. A bit of evidence to this effect is the fact that librarians do not consider "building literature" as one of their categories.

One of the problems inevitably faced, when discussing building science literature, is the organization and scope of that literature. A wealth of information in the various scientific fields related to building unquestionably exists. Each field is turning out large amounts, all having some bearing on building.

Obviously, different outlines or organizations of building literature could be developed, and most of them probably would be quite appropriate for use in this type of paper. A characteristic they would have to have in common would be the inclusion of many fields, each of which is extensive and complex within itself. Furthermore, the concept behind any of the outlines or organizations would have to encompass the interrelationships which exist between the various fields.

Figure 1 presents a meaningful illustration of the complex coverage of building science, wherein seven basic "areas," in a manner similar to that of a solar beam of light passing through a prism, are separated into a spectrum as Newton separated his seven principal colors. The seven rays making up the "principal colors" of building science might be listed as:

1) the physical and biological sciences
2) building technology
3) economics and finance
4) laws and regulations (the political and legal aspects of building knowledge)
5) basic human needs
6) design and planning
7) manufacturing and construction

*GLENN H. BEYER has been director of the Cornell Housing Research Center since 1950, and is also director of its Housing Market Survey, as well as professor of housing and development. Prior to that, he served as head of the Housing Committee of the AEC. Mr. Beyer holds an M.S. degree from George Washington University, and has written a number of books, reports and articles dealing with housing and building.

SPECTRUM OF KNOWLEDGE CONCERNING BUILDING

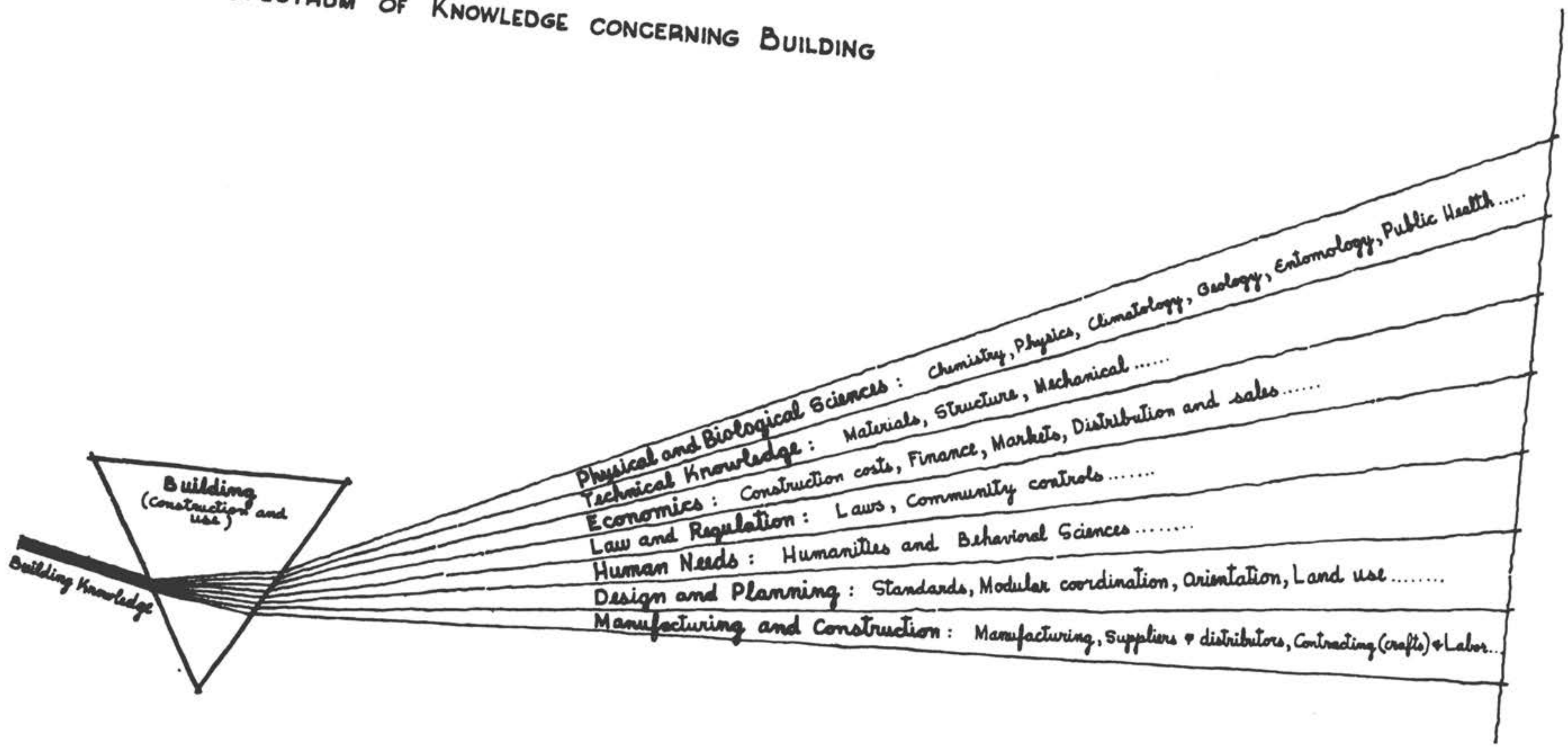

Figure 1

We might summarize these as follows: Nature, Industry, Money, Law, Man, Design and Product.

Let us look briefly at the different beams in this spectrum, one by one. First, we must depend upon the physical and biological sciences for much of our knowledge. Knowledge of the chemistry and physics of our basic materials is necessary for the invention and production of any new building materials and products. A technical report on a new chemical may show that the material has properties that fill a need in the building industry. Knowledge of soils, their characteristics and behavior in relation to structure, sanitary disposal, landscaping and planting is also needed. The importance of fungi and natural phenomena causing deterioration needs to be recognized. Climatology and weather, including temperature and humidity, wind, rainfall, etc., play their role. Closely related are the matters of insulation, acoustics, illumination, and even fire prevention.

Then, there is the whole area of building technology. In general terms, this would include the utilization of resources and knowledge of the behavior of materials, structure, etc., as well as general aspects of structural design and engineering, loads, stresses and structural performance. New types of panels, shells and floor-ceiling systems would be included here, as well as basic knowledge of building materials and products (ceramics, cement and concrete, glass, lime, gypsum and plaster, metals, etc.); also, technological information on mechanical systems and equipment (air conditioning, ventilation, heating, electricity, sanitary disposal, and communications).

A most important but quite obviously different, ray is that of economics and finance. In this area we have the matter of costs, both of construction and of operation. We have the matter of available investment funds, and the nature of those funds. We have, for housing, the whole picture of mortgage or loan financing. In addition, here, we must consider both markets for products and the broader market for the buildings themselves—and sales and distribution problems.

Then, we must recognize that there are certain laws and regulations that control, in part, the way we build and maintain our buildings. Included here at the local level are building codes, housing codes, zoning ordinances, subdivision controls and other city planning controls. At the national level, we have the Minimum Property Standards of the Government housing agencies that control design and construction, from the kinds of materials and products that go into a house, to structural design, mechanical equipment, and water supply and sewage disposal requirements.

Architectural design and planning would include primarily the translation of technical, mechanical and social data into design principles and criteria. You will note that I am not limiting the definition here to one of purely functional design. A purely functional approach to design omits recognition of other important aspects. Both technological and aesthetic objectives, for example, must accompany, but not dominate, the objective of meeting functional need. Included here would be criteria for space and dimensional requirements. Also included would be principles of design such as modular coordination, solar control and efficiency of plan. (Even the average architect today probably is able to procure within the time available to him only a fraction of what would be in the best interests of his client, the community and his own future reputation.) The many-faceted problems of city planning also would be included here, except for the legal aspects, which would fall into the category immediately above.

The sixth ray in the spectrum is that of manufacturing and construction. This area is different from building technology since it relates directly to the manufacturing and construction process, rather than focusing on the basic knowledge of materials, structures and design. Included here would be information on building materials and products (doors, windows, stairs and other auxiliary structural components) as well as knowledge of building types and of the construction process including on-site, preassembled and prefabricated construction, use of component units and other applied knowledge concerning building materials and structures, some of which would focus on cheaper production methods.

None of us is satisfied with existing materials and products; we know that there will be better processes for making them in the future. The quality of products may be improved either because we have or find a better material, or because we discover a better way of making them. As one example, the plumbing industry may find ways of combining plumbing and sanitation fixtures and fittings into package units. There may, in fact, be better methods of assembling everything that goes into a building, beginning with such simple units as the doors and windows, or more complex units such as stairs or movable partitions. Information on any developments such as these is needed by almost everyone concerned with the science of building.

Now, let us look at the ray I passed over momentarily so that I could devote more discussion to it, that of human needs. I want to devote more time to this aspect for several reasons. First, it is frequently overlooked in discussions of building technology (and some of the other areas I have already listed) and, second, it is one of the more difficult areas to discuss.

On the other hand, I believe the human aspect of building science is not only inextricably related to, but is actually basic to, all the others. I say this because I believe "good" buildings can only be built if there is an understanding of the people who are going to use them. Housing, for example, must provide a setting in which every American can enjoy the most healthy and stimulating life possible. It must contribute effectively to the development of the family, the community and the nation. Other buildings must similarly satisfy human requirements.

We would have a sterile building technology today if we did not admit that buildings are built to serve human requirements. To elaborate upon this more fully, I would like to break this discussion into two parts: first, the requirements of man as an individual which I am going to call the psycho-physiological aspects of man; and second, the requirements of groups of individuals, which I will call the socio-psychological aspects of man.

In considering the requirements of man as an individual, it becomes clear that a number of demands are made upon a building. First, it must provide for his physical health and comfort; second, it must provide for his mental and emotional satisfactions; and finally, it should satisfy his basic value orientations. Man's requirements concerning his physical health and safety have been given much more attention than the other two factors. Much has been done to provide proper thermal and acoustical environments and to provide adequate lighting, including attention to such important matters as glare and contrast. Attention has also been focused, in many time and motion studies, on the reduction of physical fatigue.

Appreciably less has been done to provide for man's mental and emotional satisfactions; in fact, to understand these still calls for much pioneering research. Yet let me emphasize, there are certain emotional factors, preferences, that often take precedence over all else. In order to get at the base of some of these, we need to begin with the human senses themselves. I doubt that the senses of taste and smell have much influence on buildings or products, so we can confine our attention to the other three, the visual, auditory and tactile.

Speaking in broad terms, the three basic physical variables with respect to buildings are space, texture and color. Related variables are temperature and humidity, light and sound. When any one of the physical variables of a building registers on one of the human senses, the result is a reaction falling somewhere along a number of continua. These continua are of the following types:

beauty ◄-------------► ugliness
simplicity ◄-----------► ornateness
satisfaction ◄---------► frustration
pleasantness ◄---------► irritation

comfortableness ◄------► uncomfortableness
spaciousness ◄--------► crowding
interest ◄-------------► boredom
relaxation ◄-----------► tenseness

To see how the process operates, let us examine the matter of lighting, for example. Here, of course, the sense of sight is involved. What are the influences of the different source-based lighting effects; that is, focal glow, luminous surround, and play of brilliance, upon people in that building? To ask the question more pointedly, where does the emotional reaction obtained from luminous surround land on the continua of pleasantness-irritation, or beauty-ugliness? The emotional response must be near the "pleasant" and the "beautiful" ends of the continua if this form of lighting is to gain general acceptability. Should not the architect and the engineer have just as great a knowledge of these factors, where any particular building is concerned, as the psychologist or sociologist?

Or, let us look at the importance of our auditory sense in buildings today. Some of our modern panels separate space quite adequately, but they hardly serve acoustical purposes adequately. One of the tests of a good building panel is the extent to which it cuts out sound. This is just another way of saying that it must fall on the "relaxation" end of the "relaxation-tenseness" continuum of emotions. Tenseness often develops from annoying noises which may not be shut out. The physics of acoustics is definitely involved here, together with the matter of insulation materials, and what they can do to solve the problem.

One final example: there are certain physical-emotional reactions influenced by the sense of touch. In the home, for example, there may be particularly unpleasant associations not only with the handling of garbage (which we can hardly be expected to resolve by changing the nature of the matter itself) but also to some degree with the cold feel of the sink and the sound of fingernails scraping on fused porcelain, metal and china. Plastics, for some people, are pleasanter to the touch than metal or even glass. The matters of materials, mechanical systems, and even sanitation are frequently involved in physical-emotional reactions.

These are some specific human problems related to building that fall within the frame of reference of building science and, therefore, of documentation of the literature in this field.

Some of the psychological aspects of man are important to building from still another point of view. The public has recently had its attention called to some new factors which are pertinent here. For example, in the last few years much attention has been given to, and much has been written about human motivations and motivation research. Recently, a book reached the top of the non-fiction best-seller lists because it seemed to provide an answer to what people's goals in life are in general. Pioneering research has shown that people hold a number of basic values which stem from their attitudes, motives, ideals and tastes. Some of this research has demonstrated that while the values are all quite different in themselves, they tend to fall into two general clusters or groupings, based on some common characteristics. On the one hand, we have such values as freedom, aesthetics, leisure and mental health. People who hold these as dominant values have been found to be more individualistic, more sensitive and more idealistic than those who do not. They are more likely to make whimsical demands, indulge in luxuries (whether they can afford them or not) and sometimes to disregard their basic physical needs in favor of luxuries and other considerations. This group is frequently motivated to buy things with an aesthetic appeal. They are sensitive to design and color, and they want their homes and even places of work planned, to the extent possible, to satisfy these basic desires.

On the other hand, people holding such values as economy, quality and physical health are more realistic and frequently less sensitive to life in general. They are more practical in their demands. They emphasize necessities rather than luxuries, and such needs as safety and physical well-being.

There is still another value, the one emphasized in the best-seller I mentioned, that of social status or prestige. The demands of persons holding this as a prime value are difficult to predict in detail since they can vary over a wide range. Individuals in this group generally have a great desire for the attention and respect of their peers. It is important to them to be represented as "the right kind of person." In some instances, they seek this distinction in sophisticated and subtle ways and in other instances through extravagant display. I am quite sure that you can recognize these qualities not only in house and office furnishings but also in buildings themselves.

This is an all too brief discussion of personal values. I would like to emphasize, however, that the factor of values is highly relevant in the field of building. Since values are derived from an individual's background, experiences, education and habits, and are solidly entrenched in most individuals, they constitute basic qualities that should not be overlooked by the architect and the engineer in the design of either buildings or products.

Now let us turn briefly to the requirements of groups of individuals, either in the home, office, factory, or some other building. These requirements in general, represent a composite of all those already discussed, plus a few others that develop when you bring two or more individuals together.

From the standpoint of building, valid criteria are needed for the space, equipment and facilities required to provide a functional environment for the group using that building. Let us take lighting as an example again, and focus on the home since the clearest example can be cited there. For festive occasions the lighting should stimulate bright, gay feelings, probably emphasizing colors, contrasts and forms. Or, at the other extreme, when only two are present, and the mood is one of romance, an intimate, even mysterious atmosphere may be desired. This is not accomplished by glowing lights and high over-all illumination; rather, it can probably be best produced by a low level of illumination, commonly achieved by weak, isolated point sources—lamps turned low, candlelight, or even firelight.

You may agree with examples of this type and still deny the importance of perception, or psychology in general, in building design and construction. Yet, the whole point of this discussion of human needs is that it is just as important to have buildings comfortable and satisfying, physically and mentally, as it is to use the right materials, have sound construction, have the mortgage funds available, meet the building codes, and perhaps even control the behavior of the termites and other pests.

At the 1953 BRAB conference it was stated that "the integration and development of this comprehensive science to a point of optimum usefulness to humanity is an endeavor which transcends the limits of any single industry, professional group, or government.... While the integration of the whole science moves slowly, the development of its many parts moves rapidly. The research literature of the various scientific fields related to building is increasing enormously. Each science, each technology, each social science discipline is turning out large amounts of research data with some bearing on the advancement of building. This wealth of new knowledge is not organized for use. It is not available in the realm of research as a tool for research programming, nor in building practice as a tool for education and progress."

Once the scope of "building science" is understood, the particular organization of subject matter is not too important. Several other factors need to be considered, however, in relation to providing better access to building science literature. Briefly, they may be listed as follows:

1) the need for clarifying the vocabulary of the numerous disciplines concerned with shelter;

2) the need for preparing and disseminating abstracts of all pertinent literature, in order to provide an over-all view of the literature appearing; and
3) the need for developing a system for the storage and retrieval of reports, articles and other types of published material needed for detailed use.

Concerning the first aspect, it is a well-established fact that every area of specialization builds up a vocabulary or jargon of its own. Anyone working in a field as broad as building constantly encounters several sets of such jargon. For example, such common and widely used terms as "value," "orientation," and "standard" have a particular meaning or definition, depending upon whether the user is a sociologist, economist, architect, or engineer. It is obvious that the lack of understanding of the intended meaning of these terms represents a formidable obstacle to the reader and user of the information concerned. It should be noted, however, that these semantic difficulties are not nearly as serious in the fields of engineering and technology as they are in the behavioral sciences.

Concerning the second aspect of the problem, a system of informative abstracts is needed in order that individuals and groups may be kept familiar with literature pertinent to their general interests, regardless of where it may appear. To illustrate this point, I have selected a few articles at random from professional journals in different fields. Generally, these articles would be read only by those with whose fields the particular journal deals. Some different fields or subject areas where particular articles may have pertinence are suggested below:

"Mass Air Conditioning for Mass Housing"

Real Estate/Building Codes and Regulations/Architectural Design and Planning Principles/Mechanical Systems/Building Operation and Maintenance/Building Types/Construction Costs

"Legal Aspects of Noise Control"

Psycho-Physiology/Real Estate/Planning Law/Urban Planning/Architectural Design and Planning Principles/Environment, Technical Aspects/Mechanical Systems/Building Codes and Regulations

"Building Types Study No. 234—Building for the Aged"

Psycho-Physiology/Sociology/Real Estate/Legislative and Administrative Programs/Planning Law/Urban Planning/Landscape Architecture and Site Planning/Architectural Design and Planning Principles/Standards/Building Technology/Environment, Technical Aspects/Mechanical Systems/Building Operation and Maintenance/Building Types

"Aesthetics and the Police Power"

Sociology/Real Estate/Planning Law/Urban Planning/Architectural Design

"The Psycho-Physiology of Skin Temperature"

Psychology/Physiology/Architectural Design/Mechanical Systems

"Changes in Attitude Toward a Low-Rent Housing Project"

Psychology/Architectural Design/Urban Planning/Sociology/Municipal Administration/Real Estate/Legislative and Administrative Programs/Building Types

"Suburbanism as a Way of Life"

Sociology/Architectural Design/Urban Planning/Real Estate/Market Analysis

If you agree that the above-listed articles would have pertinence to each of the subject areas which I have indicated (and perhaps some others), then I have convinced you of the need that exists for better access to building science literature today.

Finally, there should be a system of indexing, storage and retrieval which would permit easy access to all materials, not only those of which a user may be aware, but also those of which he may not be aware. Abstract dissemination alone will be inadequate. There will be so many abstracts that the volume will overwhelm the searcher for knowledge. Hence there must be an index to the abstracts and to the documents themselves. The coverage of such a system should include books, reports and articles of experience as well as research and experimentation. It should also include any available reports of research under way but not yet completed. (One of the problems inherent in such a task would be that of determining "cut-off" points for literature abstracted; that is, when is an item in one of the related fields not necessarily important to building science?) In this regard, the conventional library systems of cataloguing are quite inadequate.

In this discussion, I have attempted to describe the fields of knowledge covered by building science. Unlike the situation which exists in some other nations, there is a notable lack of communication in this country among many of the groups representing the different areas of building, not to mention the almost complete lack of communication between the different groups. To be sure, there are some excellent documentation and abstracting services available. Outstanding among these are services provided in such fields as chemistry, psychology and some of the fields of engineering. Currently, there are also some limited efforts in such areas as land economics, real estate, urban planning and building technology. These alone cannot begin to meet the real need for a comprehensive program of communication.

In short, the problem is that with a few exceptions, people generally do not have easy access even to the current information within their own fields, be it building technology, architecture, economics, or sociology; much less do they have access to the current information pertinent to their work in related areas. The more complex a field, the more important is the need for communication within that field.

The users of such a system of documentation would be a broad group which would include everyone undertaking research in any of the areas of building science. Such a system would also be of assistance to educators, not only teachers but also administrators. It would be helpful to planners and other public officials, whether at the federal, state or local level. Finally, a large part of the need for a sound and comprehensive system of documentation of building knowledge exists within industrial organizations.

It may be argued that there are more pressing problems involved in adequately solving all of today's building problems, problems of construction and costs or even such problems as urban decentralization and the suburban sprawl, racial bias, etc. Yet it is doubtful whether any more effective step could be taken toward their solution than to provide the powerful tools of understanding and communication in the science of building as I have defined it, tools which would enable our society to marshal and exploit all available knowledge and resources to help solve the problem at hand.

Systems for Cataloging and Retrieval of Information

By Mortimer Taube,* Eugene Miller, and Alexander Kreithen,
Documentation, Incorporated

Many forces have contrived to make it appear that there are dozens of information systems lying around or even dozens of new ones being cooked up every day so that an individual or organization with a particular problem or set of problems involving the storage and retrieval of information can take his pick or even make his own system.

We are going to defend the view that there is a science and technology of information storage and retrieval; that this technique can be applied to different subject areas; and that this application of a technique to different problems does not mean that there are diverse systems, any more than the fact that an architect may design a school or a hospital means that there are different systems of architecture, namely, school architecture and hospital architecture. Of course, an architect must understand the purposes and environment of any particular system.

Perhaps the simplest way to make this point clear is by referring to the accepted distinction between theory and practice. There is only one theory of information storage and retrieval but there are many different applications. However, the distinction between theory and practice in this field as in many others really expresses a conflict between points of view. The theorist is apt to underrate practical operating difficulties; the practical man with a special problem is apt to think that his problem is unique and may refuse to avail himself of assistance from the intellectuals.

Of course, the practical man with a problem may seek to justify his disregard of theoretical considerations by pointing out that the field of theory is a field of controversy— that he is offered not one theory but conflicting theories, each justifying different systems. I do not believe that this is the case. Nevertheless, it must be admitted that there is an enormous amount of controversy and a clamor of competing claims in this area. In large measure this clamor of competing claims can be accounted for by competition or by sheer ignorance.

The large ads for scientific personel every Sunday in the New York Times and in other media point out the fact that each company claims private know-how, each company claims that it and it alone can build the weapon, controls or system which will save America and put us into space. Whether or not we excuse Madison Avenue techniques in science, a moment's thought should convince us that each company does not have a private theory of physics, mechanics, logic, propulsion or what-not.

---

*MORTIMER TAUBE has a wide background in documentation, including service as chief of the Science and Technology Project (now ASTIA) of the Library of Congress, and as deputy chief, Technical Information Division, Atomic Energy Commission. He received his A.B. at Rutgers University and his Ph. D. at Harvard, and has also studied at University of Chicago, University of California, and University of California School of Librarianship. Dr. Taube is a member of Special Libraries Assn., American Standards Assn. and American Documentation Institute. Mr. Eugene Miller and Mr. Alexander Kreithen are associated with Dr. Taube in the firm, Documentation, Inc. Mr. Miller presented this paper at the conference and participated in the ensuing discussion.

Similarly, it cannot be true that each company or individual has a private, independent theory of information storage and retrieval which differs from the theories of all the other companies.

The fact is that among serious workers in this field there is agreement on a common body of theory, and agreement concerning those areas in which the theory must be extended and quantified. We must face the fact that we have a clamor of competing systems because so many workers in this field are in what Jacques Maritain called "that blessed state of ignorance which makes it easy for a clever man to be original." All of this is a necessary prelude to an actual concern with systems, which is the topic that has been assigned to me.

The selection or evaluation of a system to meet any set of requirements is a highly technical business requiring true professional competence. We recognize and sometimes marvel at the occasional achievement of the gifted amateur, just as those in the building industry might occasionally recognize the adequacy of a house designed and built by a gifted "do-it-yourselfer." But we also know that it is a perilous venture with not inconsiderable amounts of money at stake. More often than not, the entire budget is spent on a quite inadequate structure.

We do not know the common parameters of the building design and construction problem and therefore I cannot draw parallels for you. I can recite a few of ours, however, such as the size of the body of data, the depth of indexing, the size of the code field; the efficiency of the code; the compression noise level and redundancy of codes; the rate of search; degree of parallel search; the queuing factor; the size of vocabulary; size of terms in a vocabulary; the rate of change or obsolescence of stored information; and the size of the question. All of these and more must be determined, understood and appropriately considered in the design of a system.

We feel that it is much sounder and much more to your benefit, to say that there exists a general theory applicable to the design of an information system. We believe that many attempted distinctions among "systems" are both futile and confusing. It is far more important to be able to recognize the systems' similarities among diverse applications.

Several years ago it became clear that all information retrieval systems can be characterized by three principal attributes. First, there is the entity which is to be stored, analyzed, and ultimately delivered to the user. This may be called the item. Items may be documents, people, equipment, or any other physical entity. Second, each item may be identified with a set of descriptive qualities, subjects, keywords or phrases, or terms. These terms may vary in length from one or two words, as in Zatocoding or Uniterm analysis, to long strings of words, as in most hierarchical classification systems. Third, there is the address or location where the item is filed.

In many information systems the items have rather static properties and are of a permanent nature in the system. Systems which are concerned with scientific, technical, or other literature have the characteristic of having a collection of documents each of which remains essentially in the form as received. New items are added but those ordinarily do not automatically replace old items.

In other environments the items have an extremely transitory life in the system and usually are used to compile a set of entities which do not bear any resemblance to the items flowing through the system. Most data processing systems have this property. In many cases, the outputs are regularly scheduled and their nature highly controlled. There is usually no serious retrieval problem.

There are, however, many systems in which items have a relatively long life in the system but their nature tends to change with the input of new data. Many times there are additional terms which become associated with the items as the new data are accumulated. In addition, in such systems, the items may require rearrangement to different addresses.

We may now examine in at least a qualitative way how these notions characterize information systems. Consider first a system of microfilms of cash items in a brokerage house used primarily as records insurance in case of fire or other losses. In such a storage system the individual cash transactions are the items. The terms or subjects are usually of lesser value than the dates for recovery of the items. Thus the dates become the means of localization for address code. The items have permanence in the system and each item has one transaction.

In a typical data processing environment such as an inventory system, terms are the names of commodities, the items the stocked sizes, the addresses the bin numbers, and the transactions the changes in quantity of each size. The user is usually satisfied by a reporting of summary quantities on a regular basis.

On the other hand, there are the many document retrieval problems in which no regular method of answering the users' questions can be devised in advance. The user desires information on subjects related to his problem of the moment. Whether he knows that there is a report in the system that has exactly what he wants or whether he is merely looking for suggestions or parallelisms, he does not know in which items his answers are imbedded. The storage and retrieval problem in such an environment is a search by classes or terms for items indexed by those terms. The system must be approachable by terms with which the user can express his problem. It must deliver at least the address of all items answering a logical statement of his question expressed in the terms.

We have recently introduced a fourth basic parameter in the study of information systems. This is the concept of the transaction which is the increment of new data which changes the item in the store. It is now possible to treat all systems of storage, retrieval, and data processing as consisting of items, terms, addresses, and transactions. The operational requirements of any information system determine the relative importance of the four parameters and their quantification. From the number of items in the system, the number of terms per item, the kind of localization or address identification, the number of transactions per item and the rates of change of these parameters, the non-intellectual aspects of any information processing system can be specified.

For particular information systems environments the relative importance, quantity and rate of change of the items, transactions, terms, and addresses determine the kind of mechanization available at a given state of the art. The user load and type of output required provide the cost basis for selecting combinations of equipment yielding an optimum system.

In general, the percentage of the items in the collection which appears in the output in answer to a question influences the choice between serial, parallel or totally random access equipment. For example, in a payroll system every item or account usually is processed and its final computed transaction delivered. This obviously makes the serial approach the best choice. On the other hand, in a document retrieval system where the answers to a question may be two or three items out of a hundred thousand or more, the serial approach is contra-indicated for high user loads.

Other factors which dictate the choice of mechanization are the amount of posting of new items and their terms, the amount of computation required versus the amount of look-up, and the type of output format required.

There has been a great deal of controversy in recent years over the use of general purpose computers versus special purpose devices for various information processing problems. In the beginning large-scale computers generally were not efficiently used and consequently had excess time and capacity. Many recommendations were made to use such excess capacity for functions in which the computer was not efficient when cost was properly taken into account. Recently this excess capacity seems to be disappearing and management is looking more closely to costs and pay-off. Thus, special purpose equipment is coming into its own. In particular the problem of random access, by class questions, to large collections of items is one where little or no computation is required but memory requirements are large. Thus, there has emerged a number of retrieval devices specifically suited to provide access to items but without computation capacity.

In illustrating the unity of theory which underlies the diversity of systems it is fortunate that we have at hand a recent compilation of the National Science Foundation, "Nonconventional Technical Information Systems in Current Use." The designation "nonconventional" is not very descriptive. We can only assume that its use is based on the assumption that there is a conventional type of information system which can be specified and from which all the "systems" described in the compilation can be distinguished. We can assume further that this "conventional" system is supposedly the one used in libraries or one based on the use of Library of Congress printed cards. Actually there is no more uniformity in operating systems using 3" x 5" cards than there is in systems using 3" x 7-3/8" punched cards. On the other hand, the general theory of information systems applies equally to conventional and non-conventional systems. In fact, one of the first conclusions from the theory is that the distinction at the basis of this compilation is essentially wrong. Furthermore, a discussion of historical development of Uniterm, punched-card, aperture cards, and even magnetic tape systems out of so-called conventional systems is available in a standard library text.(1)

The NSF compilation identifies nonconventional systems as those "which embody new principles of organizing subject matter or employ automatic equipment for storage and search." What these new principles are is not specified (there actually aren't any). As for automatic equipment for storage and search, it no more establishes a difference in kind than a shift from manual to electric typewriters constitutes a theoretical change in office work.

We may assume at this point that the first distinction (namely, that between conventional and non-conventional systems) does not really establish any fundamental distinction between systems.

Within the compilation itself, systems are divided into three groups: 1) systems that store references; 2) systems that store data; 3) systems used to prepare (that is, publish) indexes. We can begin by eliminating the last group. There is a continuous literature within the library profession going back at least 75 years which deals with the conformative utility of card indexes and published indexes. Advances in technology which have occurred within the last 25 to 50 years have obliterated the distinction as of major practical concern. Using the method of shingling, tabulating or photo-listing, one can go from card indexes to book indexes as freely as one chooses. We ourselves, following a technique learned at the Library of Congress 15 years ago, have been producing Uniterm book indexes from Uniterm cards for the past six years.

We may safely conclude that a distinction between systems based upon the physical form or devices used to publish indexes is not a distinction in principle or theory and that any system of information storage and retrieval can have a published output as part of its design requirements.

We are now left with a single major distinction; namely, "systems" that store references and "systems" that store data. We have already indicated that this distinction is solely one of degree; namely, degree of specificity of the item indexed, size and

---

(1) Taube, M. and Wooster, H., Information Storage and Retrieval 1958, Columbia University Press, New York, (No. 10, Columbia University Studies in Library Service).

rigidity of vocabulary (data systems will ordinarily use smaller and more rigid vocabularies), and rate of change of information.

The compilation lists seven systems that store data:

1. The first analyzes chemical compounds considered as items under both a structural code and test data—standard IBM punches, sorters and tabulators.

2. The second does the same thing, only it is designed for a 702 computer. The statement says that the system is now being converted from a 702 to a 704. This is not considered a change in system and by the same token we should not consider a change from a complex of punches, sorters, and tabulators to a 702 a change in system. (Certainly there is a change in economics, a change in efficiency and a change in value of many of the basic variables. However, since none of these descriptions supplies any comparative data nor any awareness that they are all doing the same thing more or less efficiently, the reader of the compilation has no basis for judging whether any operative system meets its requirements any better or worse than the dozens of other designs which are possible by changing values of certain variables. As we go on to examine other descriptions we will not repeat this observation. We will presume its applicability to every "system" described in this compilation unless we indicate otherwise.)

3. The third "system" is again a structural index to chemical compounds; and almost exactly equivalent to many others described under the heading "systems that store references." It does not store nor index test data. It uses an optical matrix index of the kind designed and made available by the Bureau of Standards and the use of this device is also described in the other main section of the compilation.

4. The fourth system is practically equivalent with the first. The statement indicates that a 101 may be substituted for a standard sorter. This type of "system" change has already been carried out by many systems which are supposedly different from data storing systems.

5. The fifth system differs from the first four in subject (it is concerned with engine tests rather than chemical tests) and in the splendor of its hardware. It uses a 709, a rather expensive little gadget.

6. Here is a project which indexes properties of materials. It selects material from abstract journals. It has coded 8,300 abstracts since January 1957. It plans to use an electronic computer, a 500 card converter, punches, sorters, reproducers, verifiers and tabulators, but is at present a manual searching system.

7. The seventh system is again a chemical-biological testing system using a different make of equipment. It's nice to be told finally that one manufacturer does not have an absolute monopoly, but we are not told what the advantages or disadvantages of each manufacturer's equipment are.

This ends our concern with systems that store data. They are no different in principle from systems that store references. They are not different from one another except in subject coverage and, even with reference to subject, they are very similar.

Twenty-five different applications are listed in the NSF compilation under the heading "systems that store references" and we cannot undertake even the few sentences about each one which we devoted to the systems listed under the other headings.

Instead, I will describe them in groups. One system uses Zator cards which is a species of edge-notched cards. Eight use Uniterm Cards. One uses the Samain microfiches and his version of a "minicard" system. One uses an NBS optical matrix. One uses the IBM Special Index Analyzer. The balance use some variety or complex of data processing equipment ranging from simple punches and sorters to full-sized computers.

From reading the accounts one might just as easily change them all around, and there is nothing in any of the statements which gives any evidence that the changes would make any individual application better or worse. For ourselves, we feel that most of the applications described could be made more efficient and economical.

The purpose of this paper has been to point out how not to distinguish among systems. If we in any way have appeared to give you enough information to select a system, we have either been totally inept in our presentation or you are being somewhat presumptuous.

We think the compilers of "Nonconventional Technical Information Systems in Current Use" should make note of the theoretical unity underlying the diverse applications. Failure to do so may be performing a disservice to groups like your own. The compilation, by itself, becomes just 34 different ways of setting up and operating an information system. The presumption is that the Building Research Institute can proceed to set up any one of the 34 strictly on its own evaluation of its requirements and without study and knowledge of the basic theory of information storage and retrieval. This would be like building a house from a book of plans from a popular homeowners' magazine. Let the building industry follow its own seasoned advice by hiring a good documentation systems architect. It is no more difficult to tell a good one than a good building architect. Look into what he has actually done and decide whether you would like to have him work for you.

A Practical System for Docnmenting Building Research

By Eugene Wall,*
Engineering Service Division, Engineering Department
E.I. du Pont de Nemours & Co., Inc., Wilmington, Delaware

A number of information retrieval systems both proposed and operational have been described in the preceding paper.(1) This paper noted that there has been a great amount of discussion concerning the merits and demerits of these and other systems. All this discussion has to date generated much heat, but little light. In fact, onlookers at many conventions of documentalists might well liken those meetings to conventions of witch doctors, whereat each witch doctor can prove that in his own village his own way of curing illnesses is the only valid way. If a modern doctor listened to the arguments of the witch doctors, he would probably be able to detect in each proponent's techniques some element of medical truth, albeit well mixed with superstition. He would probably also note an underlying sameness in valid techniques among all the conference participants and would note that the differences among techniques generally were insignificant variables of superstition. Too many of us documentalists today resemble the witch doctors in the example. We cannot agree because we have no fundamental background in theory. In fact, some of us show no interest at all in fundamentals; we are more interested in the perhaps temporary success of our techniques in our own little area of interest--which we may assume to be typical. We lack understanding of environmental variables, just as the witch doctors lacked understanding of biological and psychological variables. A few of us assert that we have developed "universally applicable" techniques upon which everyone should standardize.

In this paper, however, the author wishes at least to attempt to don the headdress of an analytical, research-minded witch-doctor-documentalist, and as such, to pose a question. In this state of affairs, how can the Building Research Institute, relatively a novice in the documentation field, ever hope to winnow fact from superstition, and true general applicability from partisan pride? For there are some truths at hand in this field; there are some generally applicable fundamentals. But the word is "fundamentals"--not "techniques." To draw an analogy to military operations, it might be said that there is a fundamental strategy, but that tactics should vary with the environment. Thus, in all well-designed information retrieval systems, there are fundamental elements of sameness, despite all the sound and fury which have enveloped them.

Therefore, it would seem that the Building Research Institute would be well advised to gain an understanding of strategic fundamentals, and with this understanding it will be found possible to assess correctly the validity and values of alternate tactical

---

*EUGENE WALL has been with this company since 1943, serving as field process engineer, instrument engineer, field group supervisor (technical section) and consultant supervisor (management systems group). He earned his B.S. in chemical engineering at the University of Missouri. Mr. Wall is a member of Natl. Society of Professional Engineers and American Institute of Chemical Engineers.

---

1) Numbers in parentheses refer to the list of references at the end of the paper.

techniques. In other words, it is suggested that the fundamental problems of information storage and retrieval be defined so that the capabilities of proposed problem solutions may be compared in light of the basic characteristics of the problems. In this business, as in most others, it seems that correctly defining the problem may well turn out to be more than half the task.

Accordingly, let us proceed first to problem definition. It is believed that the definitions to be set forth are not parochial nor provincial; they are not particularly original with the author nor with those with whom he has been working. Rather, bits and pieces have been gleaned from nearly all the workers in this field and have (we think) been fitted together like a jigsaw puzzle into a meaningful, generalized entirety. After problem definition, there will be considered broad alternatives for solutions to the problems and, finally, some possible tactics which may be used to implement these generalized alternatives.

It is apparent that the problem with which we are faced is a problem in communications. Specifically, it is a problem in improving communication among three sorts of individuals or groups:

1) The originator of information--he who develops the information and he who writes it down.
2) The indexer—he who decides how the information is to be stored away so that it can later be retrieved.
3) The searcher for information--he who has a problem on which he needs help.

If it be agreed that we are faced with a communication problem, it must then be decided what sort of a communication problem. Communication problems may be of many kinds; for example, acoustical, psychological, sociological, linguistic, mechanical, etc. Obviously, we cannot hope to solve all the problems of communication, nor need we. It does appear, however, that at least two basic facets of nearly all communicative problems are significant to the Building Research Institute during its consideration of information storage and retrieval.

The first of these considerations is what may be called "feedback of information." Actually, the existence of "feedback" is a matter of degree; this is one dimension of a continuum (Fig. 1). For example, "a conversation forms a two-way communication link; there is a measure of symmetry between the parties, and messages pass

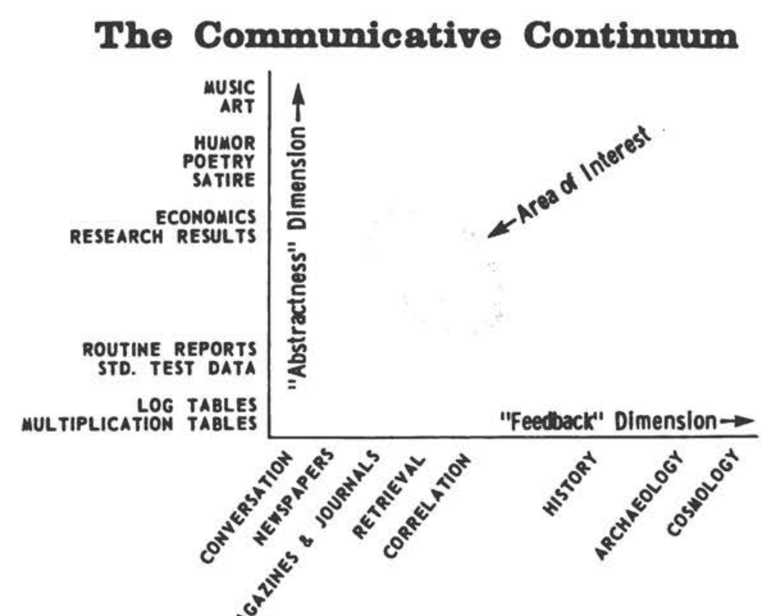

Figure 1.

to and fro. There is a continual stimulus-response; remarks call up other remarks, and the behavior of the two individuals becomes concerted, cooperative, and directed toward some (communicative) goal."(2)

Slightly farther along the continuum, the reading of a newspaper represents a unilateral, noncooperative communication (except when the reader writes letters to the editor or cancels his subscription). Still farther along the continuum, the author of a technical paper seldom has immediate opportunity to obtain feedback from his peers in the technical field in which he is interested. Of course, near the end of the continuum, there is essentially no way for the recipient to form a "cooperative link" with the originator of information. For example, an archeologist deciphering a stone inscription receives no help (except for nearby artifacts) from his forebears other than the signs carved upon the stone.

In other words, the presence or absence of feedback in the communication process, and the "process time-constant" (or "lag coefficient") of any feedback, determines to a large extent how easy it is to achieve effective communication, which is: The transmission of meaningful knowledge from one source to one or more receivers. Accordingly, this feedback dimension of the communicative continuum must, of necessity, be considered.

The other principal dimension of the communicative continuum is more familiar. It is concerned with the degree of abstractness of the information being communicated. What is meant by abstractness? Here is not meant how far the information in question is abstracted from basic considerations, in the sense that a table of logarithms is highly abstracted from a basic theory of numbers. Rather, when speaking of the "degree of abstractness" dimension of the communicative continuum, in this context it is best thought of as the degree of abstract thought required in employing the information involved. Hence, a table of logarithms would be near the low end of the abstractness dimension. By the same token, music has a very high degree of communicative abstractness, followed probably by art, humor and poetry in approximately that order.

We have here a two-dimensional communicative continuum under consideration; one dimension is degree of abstractness and the other deals with feedback. On the degree of abstractness scale, there is little practical interest, insofar as storage and retrieval are concerned, in the extreme ends (in either such things as music or log tables). Near the middle of the spectrum, however, there can be distinguished technical ideas (or information) and, near the low end, data--and we might well be interested in the part of the spectrum bounded by these limits.

Along the feedback dimension, there can be distinguished such activities as conversation, message transmission, information (or data) processing, the reporting of results of calculations, retrieval, etc. In this paper, consideration will be limited to the retrieval zone and to those matters pertinent to retrieval.

Further, because we are interested in research, our considerations will be limited principally to the information or idea portion of the abstractness spectrum, because this is the zone in which most research information falls. This is not to say that research does not develop data; rather, it is a recognition that the data developed through research usually require interpretation by words--words which stand for ideas not easily quantifiable--and this returns us to considerations of the area of communications noted in the figure. You will note that this area does not include such things as data processing, which is quite a different matter. This distinction between data and information was also made in Dr. Taube's paper.(1)

Accordingly, let us proceed to problems arising in information (or idea) retrieval. Of course, in dealing with any situation, there are three sorts of problems which arise. This is true in the realm of communications as it is in all other problem areas. First, there are the technical or intellectual problems which must be solved—and if these cannot be solved adequately, then there is no real point in worrying about any other sort of problem. However, presuming that solution of the technical problems of documentation is possible, we must also be concerned with economic problems. Technical solutions must be economically attractive—or, at least, they must not carry with them an economic penalty. Finally, there are relationship or political problems.

In line with the thesis that the technical problems must be solved in any event (and probably must be solved before the economic or political problems can be attacked effectively), let us first examine the specific technical problems in this area of communication, information retrieval. There appear to be but four technical or intellectual problems significant to this area. These are the problems of viewpoint, generics, semantics, and syntactics. The first two of these problems are characteristic of human thought, and the last two are characteristic of the particular language involved--in our case, English.

First consider the problem of viewpoint. Every individual is a unique composite of the combined, cumulative effects of his education, experience, background, environmental conditioning, and relationships with other individuals. Accordingly, individuals contemplate objects, ideas, facts or images with different viewpoints. "How you look at it" depends on how you got where you are when you are looking at it. It is not difficult to see that the word "oil" may be variously interpreted to mean: petroleum, lubricant, road surfacing material, cooking material, vehicle for medicines, fuel, source of other fuels, perfume base, hair dressing, paint vehicle, polish, etc. Many words are similar to "oil" in that it is perfectly reasonable to assign them to more than one logical class. During indexing, it is necessary to insure that variations in viewpoint among originators, indexers, and users of information will not result in missing vital information during retrieval.

The second, or generic, problem is concerned with family trees of concepts. Because each concept implies broader concepts, a literature search for information referring to broad concepts of knowledge should effectively retrieve information referring to narrower but related or subordinate concepts. (When one saws off a big branch of a tree, one normally expects all the little branches which are attached to the big branch to be removed as well--and all in one sawing operation.) For example, retrieval of all information pertaining to the chemical family "halides" should also, and automatically, result in obtaining all information on the members of that family; namely, bromides, chlorides, fluorides, and iodides.

It is normal for a concept to belong to more than one generic tree; dichlorodifluoromethane is a narrower concept within several broad concepts, such as chlorinated hydrocarbons and fluorinated hydrocarbons. In turn, chlorinated hydrocarbons and fluorinated hydrocarbons are both properly halogenated hydrocarbons.

When combinations of concepts must be considered, the family tree relationships are complicated considerably, resulting in intertwined, entangled branches which are by their very nature extremely difficult to separate from each other.

The third, or semantic, problem involves the relationships between concepts themselves and the symbols for concepts (that is, the words or terms used). Simply, the semantic problem is concerned with the relationships between words and their meanings. In this area, we become concerned with synonyms, near-synonyms, and homographs. Homographs must be distinguished from each other because they are spelled the same, but sometimes have different pronunciations and always have different meanings--for example, "flashing" (weather protection) and "flashing" (intermittent light). Other examples of semantically confusing words include base, color, lead, finish, tank, and cracking.

Another significant semantic problem is that there are situations in which two or more words have identical or very similar meanings, depending upon viewpoint. For example, within the du Pont Company, the operation of moving liquids through pipes is generally referred to as "transferring." In some cases, however, it is referred to as "transporting." If pairs of words like transferring and transporting are permitted to remain in the vocabulary of any storage and retrieval system without provision for advising searchers that the information desired may be found under more than one term, then the searcher will retrieve only that pertinent information which is included under the term he happens to use in his search. He will not retrieve that information which is listed under the synonymous or near-synonymous term. Any retrieval system must detect the situations in which more than one word or phrase may be used to describe a specific concept and make provision for cross-reference so that a searcher will be able to retrieve essentially all pertinent information on the concepts in which he is interested.

The last problem is one of syntactics. Syntax relates to the ordering or arrangements of words and the changes in meaning of a group of words which may result from modifying the relative order of words within the group. Consider "one-eyed, one-horned, flying purple people eater." This problem is particularly important in information systems which employ conceptually short terms--that is, wherein retrieval is accomplished by using terms which usually stand for single ideas or concepts. For example, coordinating the terms "fabrication" and "clamps" retrieves items which refer both to fabrication using clamps and to fabrication of clamps. Similarly, "steam" and "heating" retrieve information on the heating of steam and on heating using steam. There are also several other specific types of syntactical problems, which will be described in more detail later. These, then are the technical or intellectual problems which must be solved adequately in order that storage and retrieval of information may be effective and economical. Note that inadequate solution of the first three problems (viewpoint, generics and semantics) results generally in the loss of information during retrieval, whereas inadequate solution of the syntactical problem results in obtaining nonpertinent information during retrieval.

These definitions of the four technical problems of information storage and retrieval have proved valid during three years of quite broad experience within our company. They have also been agreed to by many other practitioners in the field. Unfortunately, there have not yet been found such useful definitions for the economic and political environments. It is suspected that most other organizations, including the Building Research Institute, may face this same situation.

Under these circumstances, it seems that there are only two basic approaches for solving the four technical problems. The first of these is, in effect, the prescription of a vocabulary for storage and retrieval. The second is the use of redundancy in storage and retrieval. Note, however, that these two basic approaches themselves constitute the extreme ends of another continuum. In practice, no system employs a precise, non-redundant vocabulary nor does any employ a completely "nonprescribed" vocabulary. Prescribed or not, the vocabulary consists of the complete set of terms used to describe the subject matter of the stored documents, as Dr. Taube pointed out in his paper.[1]

Examples of prescribed vocabularies are formal hierarchical classifications, such as the Dewey Decimal System, the Library of Congress classification, the Universal Decimal classification, and many small and local classifications. The authority lists used by many librarians are also examples of prescribed vocabularies.

Please note that here the term "classification" is not used as a synonym for "shorthand description" or the like; rather, it is used in the restricted sense of formal hierarchical arrangement, wherein ideas are included as subclasses of broader ideas, etc., and are formally arranged, for retrieval purposes, in such a manner.

In light of the previous discussion on the abstractness dimension of the communicative continuum, it seems apparent that a prescribed vocabulary (such as a formal hierarchical classification) operates most effectively at the low abstractness end of the continuum and becomes less and less effective as the highly abstract portion of the continuum is approached. Here, meaning does not reside within the word or symbol alone but rather depends principally upon viewpoint. Accordingly, it would be expected that prescribed vocabularies would be used for data retrieval systems, but not for information (or idea) retrieval systems. And, because the Building Research Institute is concerned with research information, which ordinarily must carry prose along with any numbers (that is, data) in order to make the numbers meaningful, the Building Research Institute must also be concerned with a somewhat abstract portion of the communicative continuum. Thus, prescribed vocabularies would not be generally applicable to Building Research information systems.

It is apparent that prescribed vocabularies, such as formal hierarchical classifications or authority lists, will be advantageous if one or more of the three following situations prevail, as they always do in well-designed data retrieval systems:

1) The collection of documents is small, so that they do not need to be subcategorized too greatly.
2) The field of technology covered by the stored documents is narrow, so that the prescribed vocabulary can be small.
3) The number of potential users of the stored information is small, so that the conventions necessary in using the prescribed vocabulary may be policed effectively.

Please note that we are not objecting to formal classifications when they are used for organizing one's thoughts or in getting an over-all view of a total situation. We are merely saying that the utility of a classification is much less in information storage and retrieval than it is under those other circumstances.

Can the other alternative—redundancy—be employed in storing and retrieving information or ideas? Certainly, to many people, redundancy is a nasty word. To them, it implies something unnecessary, repetitive, verbose, or the like. To the communications engineer, however, redundancy is something quite different. While it may still be repetitive in a certain sense, redundancy is essential in order to insure that a signal is not lost in the noise which may exist in the communication channel.

Even in everyday conversation, we are unconsciously using a great deal of redundancy. It has been said that the English language is more than 50% redundant and that if someone could speak with no redundancy we would be unable to understand what he said. So, redundancy does have its virtues; what we have come to think of as being undesirable is excessive redundancy. And, "excessive" depends upon the circumstances, upon the uses to which we wish to put this principle of redundancy. In some instances, it would be advantageous to employ redundancy, whereas in other instances, it would be avoided insofar as possible.

There are two ways in which redundancy may be profitably employed in the storage and retrieval of information. The first is that one may index redundantly; that is, one may index under all probable viewpoints, at all probable levels of generality of viewpoint, and with all probable terminology which may be employed by the originators, indexers and users of a system. Alternatively, however, one may index by taking into account only immediately apparent viewpoints, generics and semantics, but we may search redundantly; that is, by translating each individual inquiry into a number of different inquiries, using search terms standing for probable viewpoints, probable levels of generality of viewpoint, and probable terminology. This means that each individual inquiry will result in a number of questions—each question being composed of a permutation of different terms which might have been used to index the desired information.

The choice of whether redundancy is employed at the input or at the output end of a system depends purely upon economics. It depends especially upon the search cost per question, which depends in turn upon the number of searches requested per unit time. Against this must be balanced the accession rate of new documents into the system and the unit input cost for redundant indexing. For systems which hold a large collection, especially systems which have a high ratio of references to accessions, the designers should consider carefully the desirability of employing redundancy at the input side. In either situation, the use of redundancy results in a "continuous" solution to the problem of cost versus effectiveness. That is to say, if cost is plotted against effectiveness of retrieval, a smooth curve will result (Fig. 2). Costs will be

Economics of Retrieval

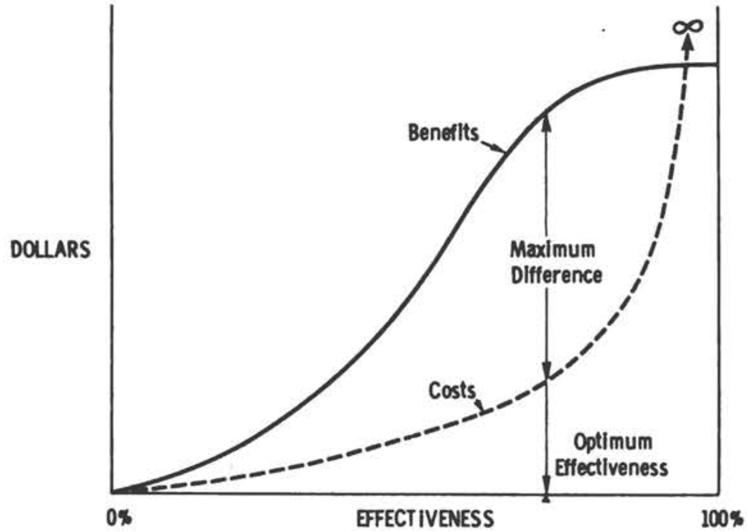

Figure 2.

zero at zero effectiveness and will increase with a steadily increasing slope, approaching infinity as 100% effectiveness is approached. There will be no discontinuities on the curve, such as often occur when attempting to employ a prescribed vocabulary; for example, when a classification must be entirely redone in order to keep up with changes in technology.

It is believed that the relationship between benefits and effectiveness is "S-shaped;" that is, benefits rise slowly as effectiveness is first increased from zero and then more and more rapidly, yet the curve finally flattens out as 100% effectiveness is approached. In practice, one would wish to operate at the effectiveness level which provides the greatest difference between benefits and costs.

At present, unfortunately, the choice of this optimum level of effectiveness must be largely a subjective one, because benefits cannot yet be quantified well. The desirability of the "continuous" solution to the technical problems, however, must not be minimized, because it provides a true capability of adjusting storage and retrieval operations to any economic facts of life which may develop at a later date. This is certainly one significant advantage, among many, of using redundancy rather than vocabulary prescription to solve the viewpoint, generic and semantic problems.

If it is presumed, based on this reasoning, that redundancy rather than vocabulary prescription is probably the most advantageous principle to employ for storage and retrieval of information, there is then created another problem. How can this redundancy be obtained? If one elects to employ redundant indexing, where will one find this paragon of an indexer who can use all probable viewpoints, generics and terminology? If one elects to employ redundant searching, where will one find a person who can compose inquiries employing all the permutations of probable viewpoints, probable generics, and probable terminology?

To a minor degree this problem has already been solved, for there is in existence a device known as Roget's Thesaurus. Unfortunately, this thesaurus includes many terms in which we are not interested and excludes many technical terms in which we are vitally interested. As it now stands, its usefulness is limited. However, the principle is still valid. An appropriate technical thesaurus would serve as a "word-guide list," a "word-reminder list," for the indexers and retrievers of information. Such a thesaurus would indicate synonymous terms, generic relationships among terms, and other relationships among terms. The possibility of using a thesaurus for solving the viewpoint, generic and semantic problems is not original with our company. Its advantages have already been described by several others, including Bernier (3) of Chemical Abstracts, Heumann (3) of the National Academy of Sciences-National Research Council, Luhn (4) of IBM, and Taube (5) of Documentation, Inc.

The creation of such a technical thesaurus is not merely a theoretical possibility. In our engineering department, we have constructed a thesaurus for use in information storage and retrieval (Fig. 3). We have found it essential for achieving system effectiveness.

When first considering the creation of a thesaurus, one is inclined to wonder whether all the hundreds of thousands of words in the English language must be included, and if so, one is appalled by the magnitude of the task. But, in fact, the vocabulary of science is quite limited. Numerous investigators have pointed out that the vocabulary of any one field of technology is limited to approximately 5,000 terms, that the vocabulary of all technologies is limited to approximately 20,000 terms, and that the whole of human knowledge could be expressed in less than 40,000 terms. These vocabularies are, of course, descriptive vocabularies; they do not include names of people, places or things— nor even such things as names of chemical compounds, for which there appears to be no growth limit. This category of terms known as names, however, is one which causes only minor difficulties, in practice, in the operation of information systems.

Presuming existence of a thesaurus, how might it be used? If redundant indexing is to be employed, the indexer might first list as indexing terms those words or phrases which are used by an author to describe the information he is attempting to communicate. The indexer could add words or phrases of his own further to describe the information in the document at hand. He could then refer to the thesaurus to obtain generic and other terms related to those terms already listed as index entries. Depending upon the value of the information contained in the document, the indexer could use the thesaurus to whatever extent might be appropriate. This determination would be a subjective one. The indexer must obviously be competent to a considerable degree in the field of knowledge which he is indexing.

In this manner, the indexer could describe information from all probable viewpoints, all probable levels of generality of viewpoint, and with all probable terminology. Of course, the thesaurus would have to be a living, growing document, one which would be subject to continual modification and updating. Such modifications, however, could be made easily on a piecemeal basis; this contrasts with the wholesale modification which must be made periodically to formal classifications and the like. The thesaurus approach carries with it one possible penalty. It forces one to employ conceptually "short" indexing terms. Here, "short terms" means unit concept terms, as distinguished from terms standing for combinations of concepts. The use of conceptually "short" terms is necessary because we have found it impossible to construct a thesaurus in which the terms stand for combinations of concepts. The same inherent, logical factors have forced every lexicographer, from Samuel Johnson on, to use terms which stand essentially for unit concepts in all dictionaries and thesauri ever created. The reason is, of course, that terms standing for combinations of concepts (such as phrases, Dewey Decimal numbers, or the like) are so specific in meaning that they can hardly be defined in terms other than themselves. For example, what single term could stand for a single concept which would be generic (in the inclusive sense) to the combination of concepts expressed by the phrase "evaluation of foamed plastic insulation under variable climatic conditions."

EXAMPLE OF A TECHNICAL THESAURUS

383100 S. AND R. DIVISION-SEE SALVAGE & RECLAMATION DIVISION

432700 S.B.B. ACID-SEE SULFOBENZOYLBENZOIC ACID

417700 S.P.-SEE SPLAY POINT

381700 SABINE RIVER
PO 378300 RIVERS
PO 482100 WATERWAYS

381900 SADDLES/PACKINGS/
GT 44900 BERL SADDLES
RT 80300 CERAMICS
RT 316400 PACKINGS
RT 363800 RASCHIG RINGS
RT 392200 SEPARATION

382000 SADDLES/SUPPORTS/
RT 111800 CRADLES
RT 201200 FOUNDATIONS
RT 316700 PADS
RT 436000 SUPPORTS

382100 SAFETY
RT 1600 ACCIDENTS
RT 61000 BURNS/INJURY/
RT 101300 CONDITIONING
RT 191000 FLAMEPROOFING
RT 209700 GLASS
RT 219600 HAZARDS
RT 220100 HEALTH
RT 239400 INJURIES
RT 331400 PHYSIOLOGICAL
RT 350900 PREVENTION
RT 355900 PROTECTION
RT 356400 PSYCHOLOGICAL
RT 383600 SANITATION

382300 SAGGING
RT 42300 BENDING
RT 112800 CREEP
RT 124100 DEFLECTING
RT 129400 DEPRESSING
RT 393000 SETTLING

19200 SAL AMMONIAC-SEE AMMONIUM CHLORIDE

382400 SALARIES
PO 320700 PAYMENTS
RT 98700 COMPENSATION
RT 320800 PAYROLLS
RT 480200 WAGES

Figure 3.

One is thus forced into the use of conceptually "short" indexing terms—often single words and almost invariably terms which stand for unit concepts. This is not necessarily bad; there are actually some distinct advantages. The principal one is that an enormous body of knowledge can be described with relatively few terms, just as we can compose an almost unlimited number of English words using only 26 letters. In order to do this, however, our terms must be combined at the output side of the system rather than at the input side. In essence, such a technique consists of finding the information which exists at the logical intersection of terms and, accordingly, systems which employ such techniques are known as "coordinate" or "concept coordination" systems. However, information signified by the logical union of terms can also be retrieved in such systems.

Another advantage of using unit concept terms is the system simplicity which results; the complexity of information storage and retrieval systems depends largely upon term length. Short-term systems, besides being simpler, also tend to be less bulky and easier to use.

The use of unit concept terms, however, does carry with it one rather serious disadvantage. The syntactical problem is intensified. Coordination of the terms "cooling" and "water" will result in retrieval not only of the documents dealing with cooling water but also of those dealing with water cooling, an entirely different idea. In other words, the use of unit concept terms intensifies the "noise" problem in the communication channel; the tendency will be to retrieve more nonpertinent information than if the terms were longer. Note however, that the system at least "fails safe;" the probably pertinent information is retrieved, but unfortunately an amount of nonpertinent information is also retrieved. This may, in large systems, be quite undesirable.

What, then, is the solution to the last of the four technical problems—to the syntactical problem? Again, an empirical approach is suggested, an approach which has as its goal adequacy rather than perfection. Let us construct a logical model of an index such as has been described. This model will be a binary rectangular matrix (Fig. 4). Each horizontal row of the matrix will stand for an item of information stored in the system. Items may be journal articles, books, reports, file folders or the like. Each of the vertical columns of the matrix will stand for an indexing term, with each such term symbolizing a unit concept. Then, to distinguish the terms by which any particular item is indexed, place a "1" at the appropriate term-item intersections. All intersections which are not occupied by a "1" may be presumed to be occupied by a "0."

In this simplified example, Item 1, as indicated by the topmost horizontal row of the matrix, discusses "cooling water." Item 2, the middle horizontal row of the matrix, discusses "water cooling." Item 3 discusses "cooling of air with water."

Thus, in our hypothetical matrix, there are three terms. These will be "water," "cooling", and "air." We will not be concerned here with the viewpoint, generic and semantic problems--only with the syntactical problem. Note that if all information on "cooling water," is desired, the inquirer will be referred to all three items, whereas only Items 1 and 3 are directly pertinent.

This simple problem can be solved by an oversimplified use of what we have called "role indicators" (Fig. 5). The two materials terms, "water" and "air," are divided into "use of" and "passively receiving an action" portions. Now the "noise" in the simple example has been eliminated. It can be seen that Item 1 discusses "the use of water for cooling," that Item 2 discusses "cooling of water," and that Item 3 discusses "the use of water for the cooling of air."

Of course, two role indicators are insufficient for effective storage and retrieval. We have found it necessary to subdivide our terms by as many as 12 roles; others may find it desirable to develop fewer, more, or different roles. If role indicators are carefully designed, as we believe ours are, they may be quite generally applicable to numerous broad fields of knowledge. In fact, several of our roles carry grammatical

"Model" of Index

TERMS →	AIR	COOLING	WATER
ITEMS ↓			
1	0	1	1
2	0	1	1
3	1	1	1

Figure 4.

Modified "Model" of Index

TERMS →	AIR		COOLING	WATER	
ROLES →	USE OF	PASSIVE		USE OF	PASSIVE
ITEMS ↓					
1	0	0	1	1	0
2	0	0	1	0	1
3	0	1	1	1	0

Figure 5.

connotations such that we are, in effect, inflecting our terms in order to avoid syntactical ambiguity. Please note, however, that we are not attempting to provide precise grammar in our indexes. Rather, we are only attempting to make it possible to provide sufficient grammar so that the "noise" may be minimized. A term-with-role-assigned is essentially a precoordination of the term with an implied definitive concept term which imparts to the term-plus-role an element of syntax or word-ordering so that stored information produces fewer false associations. This means that there are three basic requirements which must be met by a set of role indicators:

1) They must be indicative of broad concepts which are encountered very frequently in the particular environment of the information system;
2) They must, insofar as possible, be nonambiguous among themselves (that is, mutually exclusive) and, accordingly,
3) They must be few in number.

Even the use of role indicators does not completely solve the syntactical problem. Role indicators alone will not prevent "noise" when a document discusses, for example, the corrosion of iron in sulfuric acid and the corrosion of copper in nitric acid. Role indicators would not prevent the retrieval of the document when information on the corrosion of iron in nitric acid is requested. This problem, however, is easily solved by subdividing the horizontal item rows in the matrix in a fashion similar to that in which the term columns were subdivided. This, in effect, means subdividing the physical items of information into smaller items. This is best done by an intellectual subsectioning, not a physical or "geographical" subsectioning. One might say that we really end up by preparing individual sentences describing the document at hand, with each sentence having its elements made more precise by the use of role indicators.

In such a manner, the original binary rectangular matrix is made into a much finer-grained matrix, but it still remains a binary rectangular matrix and is logically equivalent to the system described earlier. By subdividing terms and items, we have not eliminated all "noise," but are only permitted to make a reasonable compromise between "noise" elimination and effectiveness of retrieval.

At this point there may be discussed some practicable tactics in the implementation of the more basic considerations of storage and retrieval of information. There is now a choice as to how the terms and items may be grouped. Consider again the binary rectangular matrix, with or without terms and items subdivided, because it will make no difference, the matrices are logically the same. Because in the actual physical world it is impracticable to arrange everything into a huge binary rectangular matrix, the information must be grouped either according to items (the horizontal matrix rows) or according to terms (the vertical matrix columns).

Let us call systems grouped according to items (the horizontal rows) randomly grouped systems (Fig. 6). An example of such a system would be one in which a card-fed computer of some sort is employed for retrieval purposes; each card (Fig. 7) would be

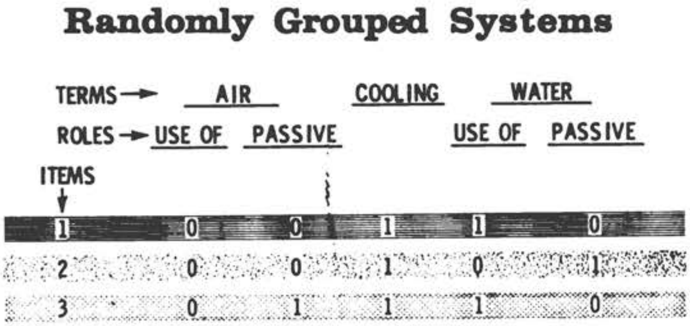

Figure 6.

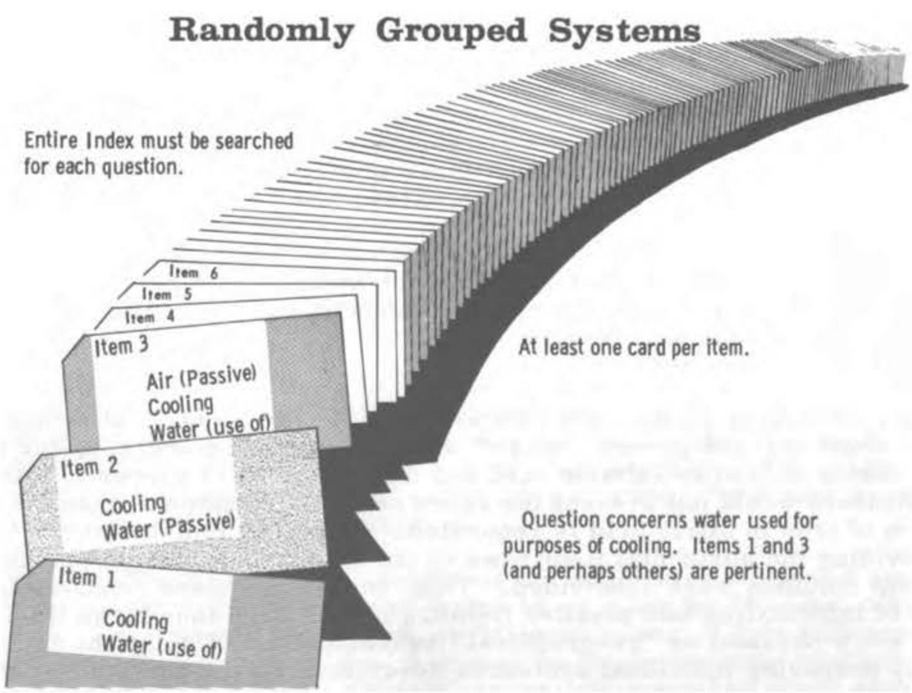

Figure 7.

occupied by the index entries for one given item; on that card would be grouped together the item number; that is, the "address" of the actual physical document, and all the index entries for the item. The following card would contain the next item number and the index entries for that item, etc. Items would be entered on cards as they arrive at the information system and are indexed. Thus, the most recently arrived item would occupy the last card in the system. Hence, this would be called a randomly grouped system because the subject matter would have no effect upon the arrangement of the items in the store. When searching such an index, it is necessary to examine each of the items, one by one, from the first item to the last. Such a searching method forces one to look at a tremendous amount of information in which one is not at all interested. You may say that the randomly grouped store could be subdivided into various subject classes, but the problems involved in classification have already been discussed, and such action does not appear to be indicated for large systems.

Let us call systems grouped according to terms (the vertical columns) pre-filed systems (Fig. 8). In such systems, one might have a card standing for each term in one's vocabulary. On this card, one would enter the item numbers, that is, the document "addresses", of the items which have been indexed by that term. Now, if one wishes to retrieve information, one has only to examine the portions of the store which are most likely to contain the information in which one is interested (Fig. 9). That is, of the total set of cards, examine only those few cards standing for the concepts pertinent to the question--in this instance, the cards for "cooling" and "water, use of." Note that there are three pertinent references. This contrasts with the searching method in randomly grouped systems, which involves examining each of the items, one by one, from the first item to the last.

Both the randomly grouped and the pre-filed systems are, of course, logically equivalent and absolutely equal in retrieval power and effectiveness. They may, however, have significantly different characteristics insofar as economics are concerned. Especially for idea or information systems and particularly where the collection size is large, the pre-filed method appears to be best. The choice of grouping method, however, is not fundamental, because a system grouped in either fashion can be changed to the other fashion at any time with no loss in effectiveness; there will, of course, be a one-time economic penalty in the change-over operation.

Figure 9 illustrates one form of pre-filed index--a centralized but simple card file. Figure 10 is an example of the same basic sort of index which illustrates provision for decentralized searching. This index format is designed to permit publication and wide dissemination. The entire index is duplicated, side-by-side, in one book and the two sides are bound independently at the top; this facilitates the easy comparison of item numbers listed under any two terms. In addition, the item numbers listed under each term are divided into ten columns according to their terminal digits and this, too, facilitates searching.

In the example shown, the terms "Catalysts-Role 11" and "Concentration (Composition)--Role 2" express the question, "effects of catalyst concentration." When these terms are coordinated, it is seen that there are 17 pertinent documents, as indicated by the item numbers which are encircled under each term. These matching numbers can now be matched against numbers listed under a third term, and so on. Such a simple index, updated periodically, will usually be a satisfactory retrieval tool until the document collection becomes quite large.

At this point, it may be asked: "How might some of these tactics be implemented in practice?" Below is detailed an example of an indexing and retrieval system based upon the principles already described. It must be understood that there are many possible and legitimate variations upon this major theme. With reservations concerning the lack of general applicability of the tactics, the example can be described as follows:

Following selection of a document for inclusion in the collection, the following steps may be taken:

1) An accession number, or "address," is assigned to the document; this is usually the number next higher than the last previously assigned accession number.
2) A technically qualified person indexes the document by doing the following:
   a) He analyzes the document to determine its information content. This analysis step is the same for all retrieval systems; there is no short-cut for having a qualified person gain an understanding of the content of the document. This step is also the most expensive step in the input procedure. Of course, the cost of document analysis may be controlled by various policy decisions. For example, the policy may be to skim through the document

Pre-filed Systems

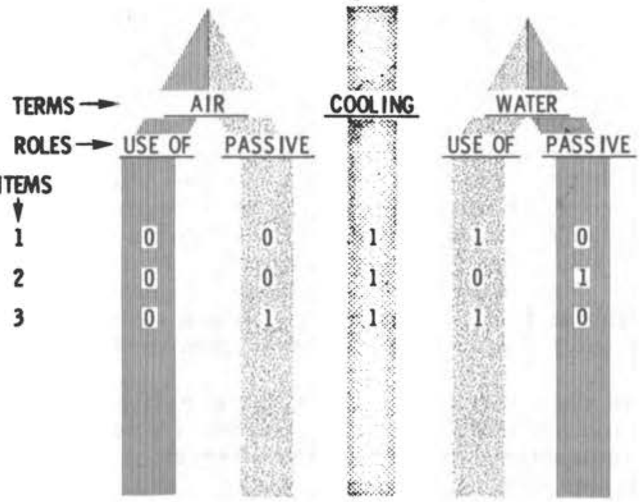

Figure 8.

Pre-filed Systems

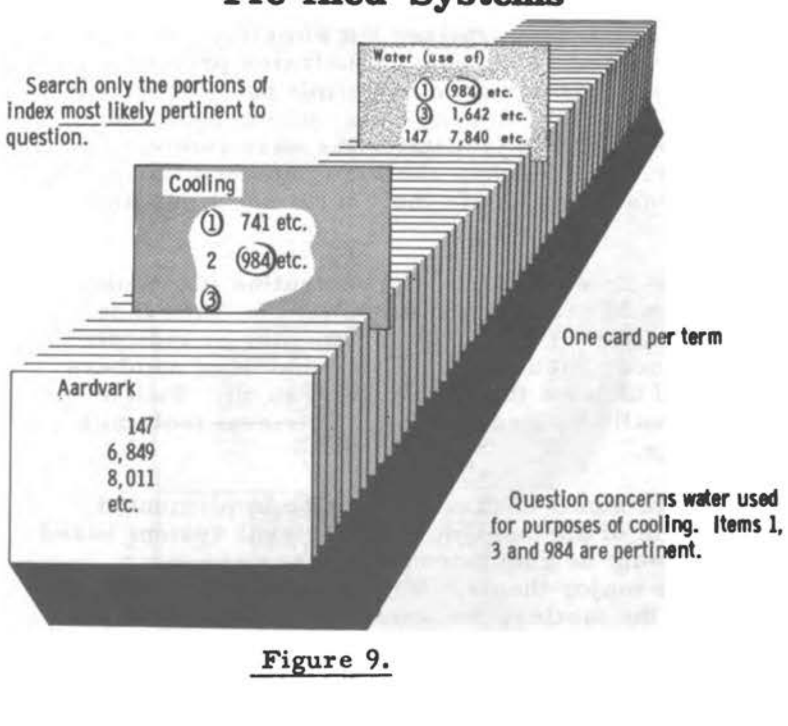

Figure 9.

EXAMPLE OF A PRINTED INDEX

CATALYSTS (1 - use of)

70A 70E 90A 100A 100B 110A 120A 130A 150D 220B 220C	11A 11B 21B 21D 71A 71E 71F 71G 81A 81B 91A 121A	12A 22A 22J 72A 82B 82F 142A 142C 212B 212C 222E 232A 232L	13A 13B 43B 73B 73D 83A 113A 153A 173A 193A 203A 203H 213A 213C 223A 243A 243B 263A 263B	4A 24B 34B 34C 34H 44I 74A 74N 94A 114A 164A 204A 214B 214C	35A 35D 45D 45E 75A 215D	6A 26A 26B 36A 36B 86A 106J 136A 206A 266A 266K	17B 57A 57C 67A 67B 67D 87A 87J 107A 107H 147C 147D	8C 8D 28F 88H 108C 118A 158D 248A	19B 19D 79A 79G 79H 109A 129A 149C 189H 189G 219D

CATALYSTS (2 - causes)

90D		82D 142A	83B 173B 263A	34I 114D 164D	75C 75D 125B	86B	17C	28B 28G 68E 88H 108E	19D 19F

CATALYSTS (3 - reactants)

			223A						19G

CATALYSTS (6 - Contaminant)

			153B						

CATALYSTS (8 - research on)

110A 130A	71F	72A	83A						

CATALYSTS (9 - effects)

									79L

CATALYSTS (11 - acted upon)

110E 120A 130B 260G	21A 71E	72A 82D 142B 142D 182A 232L	13C 13E 73B 83A	14K 74I 114D 164D	25C 75C 125B 195F	26D 26G 66A 86B	17C 17F 67B 87E 87F 107D 107E 107F	88H 158D 248E	19F 79J 79L 79N 109H

CATIONIC (adjective)

			123A		35A 35D 75A 75B 75E				

CONCENTRATES (10 - design of)

	271B								

CONCENTRATES (11 - acted upon)

	271B					166A 236D			

CONCENTRATION (COMPOSITION) (2 - causes)

10B 10F 60E 60F 90A 90B 120A 260D 260E 260G	31E 101A 111F 111G 111H	82D 122C 122G 232G 252C 252D 272A 272C 272D	83B 123D 123G 203A 233E 253D 263A	14I 14J 14K 114D 164D	15F 65F 75C 115A 125B	26D 46H 76C 76D 86B 96B 266D	17C 47H 67B 67E 87F 107D 197B 237C 237D 237E 237G 237I 247D	28B 88G 88H 208B 248E	49B 69C 79M 219C 229G

CONCENTRATION (COMPOSITION) (8 - research on)

30K		122B 122L 232H	233I	54D		106E		228B	229E

CONCENTRATION (COMPOSITON) (9 - effects)

90B	111F 241G	122C 222D	123A 123B 233D 233H		145D	76C	77D 177A	228B 248E	99A 219G

CONCENTRATION (COMPOSITION) (11 - acted upon)

			203E 233D	264E	255D		77C	248C	

CONCENTRATION (PROCESS) (1 - use of)

				64D					

CONCENTRATION (PROCESS) (8 - research on)

			153B 233D 233H	64A		16D 116C	117A 127F		

CONCENTRATION (PROCESS) (9 - effects)

			233H	64B 64C			117A		

CONDENSATION (PROCESS) (1 - use of)

130C 150A	41A 41B 101A		33A	24E 44B 44C 44E	65B 85A 85B 85D 185A 185B	6C 16D 76A 106E	47C	88Q	149A 189D

Figure 10.

and not by any means to understand completely its content. This will make for relatively shallow indexing and lower costs. On the other hand, the policy may be to have the indexer understand and index the document comprehensively, which makes for higher costs.

b) The indexer identifies the concepts of knowledge discussed in the document. This step may be performed along with the analysis step.
c) The indexer evaluates for importance the concepts of knowledge discussed in the document; he chooses certain ideas for indexing and discards others.
d) The indexer describes the information content of the document; this is the physical process of indexing. In essence, this is the thinking out of a set of declarative sentences. The important ideas in these sentences are used as indexing terms. Appropriate role indicators are used to indicate the relationships among these terms. Thus each such sentence constitutes in the index one subdivision of the document, as described earlier in re-subdividing the horizontal rows of the matrix.
e) The indexer then adds other terms, to the extent justified, using as source material the thesaurus listings under some or all of the terms he has already extracted from the document. Assignment of role indicators to these additional, redundant terms is routine, being determined largely by the role indicators assigned to the originally developed set of terms. The cost of input can be controlled to some extent, during this step, by making policy decisions concerning the extent of use of the thesaurus.

This is the end of the intellectual work of input. It is also the end of the most costly part of the operation. Next, the indexing entries (both terms and accession numbers) for a number of incoming documents are sorted (clerically or mechanically) into the pre-filed order and are posted (clerically or mechanically) to the index itself in whatever form it may exist, either manual or mechanized.

Searching techniques, of course, depend largely upon the physical form of the index. Manual card or printed indexes often depend upon the manual matching, the coordination, of lists of item numbers posted on terms. Machines do essentially the same thing, but can often be programmed to make complex searches in one rather than in many steps.

In fact, mechanization of an index is merely another tactical problem. It is apparent that the degree of mechanization will depend upon a number of environmental factors, such as the size of the index, the method of grouping of the index, the number of questions which have to be answered per unit time, how soon a question put to the index must be answered, and other factors. If the total store is small, if only a few questions come into it each year, and if there is no particular urgency in answering these questions, obviously a very simple manual system would be the appropriate choice. Other circumstances, however, might necessitate an extreme degree of mechanization.

At this time, one point must be emphasized. Almost any machine or device designed for manipulating data or information can be employed in the storage and retrieval operation. They can all be programmed or wired to do the same job. Naturally, there may be differences in the way that one has to arrange one's system in order to use a given machine. For example, some machines would require that one arrange the file in random order. Others might dictate arrangement in a pre-filed order. Such considerations might have a major bearing upon the economics of mechanization. And this is the only basis upon which mechanization should be considered—that of economics.

Another tactical question to be resolved is this one: "After the searcher has found (in the index) the identification numbers of presumably pertinent documents, what then does he do?" It has been the experience of most of us that the searcher should at this point have available a set of some sort of abstracts. He can look up the abstracts of those documents to which he is referred and decide quickly which ones are most important to him, which ones to obtain and to read first, which ones to defer action on, and perhaps which ones to ignore altogether.

On the other hand, abstracts without an index serve only a limited purpose, that of advising the reader of current events, which may or may not be of interest to him at the moment and which he may not remember at some future time when they should be of interest. Further, the scanning of large numbers of abstracts is laborious and time-consuming; an index to the abstracts permits much higher search efficiency.

In summary, it may be said that the details, the mechanics, of an information storage and retrieval system are not too important at the stage of the game at which the Building Research Institute finds itself. Rather, basic considerations and the building of a firm foundation should be paramount. It is believed that what has been described herein is fundamentally correct, insofar as basic considerations go. As we proceeded more into tactics, however, there may well be some honest disagreements. It must be reiterated, however, that the system described today does provide a solution to the storage and retrieval problem. It permits the operators of the system to obtain greater system effectiveness, better retrieval, if they are willing to pay more for that greater effectiveness. It does not provide perfection, but who has ever felt rich enough to build a bridge, or to design a dam, which he can guarantee to be perfect for all time? Who has ever found that which he is willing to guarantee will be the least costly solution to any problem for all time? We can only state a firm belief that an information system built upon the principles discussed will be capable of meeting one's needs to the extent that one is willing to pay for that capability and, should more economical techniques be developed in the future, that the system may be converted at small cost.

The last statement is based upon mathematical considerations of the binary rectangular matrix which we discussed earlier. These considerations have led, we believe, to the beginnings of a mathematical theory of written communication—a theory which employs the same equations and leads to the same end-results as those developed by workers in other fields. This, we think, attests to the fundamental nature of the considerations discussed, and indicates that while tactics may change because of different environments, the strategy of a well-designed system will remain valid.

REFERENCES

1) Taube, Mortimer, Miller, Eugene and Kreithen, Alexander, Systems for cataloging and retrieval of information, as presented at BRI Fall Conference November 1959.
2) Cherry, Colin, On human communication 1957, John Wiley and Sons, New York.
3) Bernier, Charles L., and Heumann, Karl F., Correlative indexes III, Semantic relations among semantemes, The Technical Thesaurus, American Documentation Vol. VIII, No. 3, pp. 211-220 1957, Interscience Publishers.
4) Luhn, H.P., A statistical approach to mechanized encoding and searching of literary information, IBM Journal of Research and Development Vol. I, No. 4, pp. 309-317 Oct. 1957.
5) Taube, Mortimer and Associates, Studies in coordinate indexing, Volume II, pp. 72-111 1954, Documentation, Inc., Washington, D.C.

Panel Discussion

Moderator - C.H. Topping, Chairman,
Senior Architectural and Civil Consultant,
E.I. du Pont de Nemours & Co., Inc.

Panel Members - Glenn H. Beyer, Director, Housing Research Center, Cornell University
Julius Frome, Patent Research Specialist, U.S. Patent Office
C.D. Gull, Information Systems Analyst, General Electric Co.
Laurence B. Heilprin, Physicist, Council on Library Resources
K.F. Heumann, Director, Office of Documentation, NAS-NRC
Eugene Miller, Documentation, Inc.
J.S. Sanford, U.S. Dept. of Defense
F.R. Whaley, Head, Tech. Information Services, Linde Company Div. of Union Carbide Corporation

Mr. Topping: We would like to throw some light this afternoon on what's going on now in documentation that would be available and valuable to the building industry, and what is planned. There is a great necessity, as has been pointed out, for assistance in collecting and disseminating, coordinating and making available the tremendous amount of information already existing on building science. There is a great deal more to be known about building science than we now know, and we think perhaps the wisest thing for us to do is to attempt at the start to document principally new research, current research, on building science. This is the job we think we can get our hands on and get done. We may expand that later, if we have the funds and the system works and everything is favorable, to include other things than new building research. We do not expect to decide this afternoon, obviously, that the Building Research Institute is going to do this, but we hope that information will be brought out that will enable the Research Committee and ultimately the Board of Governers of the Institute to decide whether this is feasible and whether we should do it. So, I would like to ask the panel, "Is the proposed method that was discussed this morning soundly based generally for our purpose?

Mr. Frome: I think that the fundamental principles enunciated by Mr. Wall this morning are excellent for the following reasons. They would enable the Building Research Institute to establish a good foundation for a system which can start off in a small way and grow to be a large one, without fear of being constrained or not being flexible enough. One of the features of the system which I feel is excellent is the establishment of the roles or links. This adds something which for a long time has been lacking in some of the coordinate indexing schemes. Second, because of the thesaurus approach, the dictionary is flexible and can be enlarged easily. We in the Patent Office have followed along similar lines and we feel that it will eventually help solve our problem. The techniques used are susceptible to both manual operation at the inception and mechanization when the file grows rather large.

Mr. Topping: Do you think that this system, then, is suitable for the documentation of building research, specifically?

Mr. Sanford: I thought Mr. Wall's paper this morning was an excellent analysis of the basic problems which face anyone about to embark on a sophisticated information system. It is apparent that you gentlemen would profit mightily from an information system which has flexibility, essential economy of operation, and the potential for growth. One of the requirements of an information system for the Building Research Institute, it is clear from what I have been reading in the past days about your problems, is that it must be many things to many people. You represent wide diversity of interest; you have to satisfy a wide variety of customers in a wide variety of ways, and thanks to the work that has been done in this field since World War II, the art and science have now developed to the point where flexibility and economy of operation of this sort are possible. Certainly one of the things which has bedeviled your field in Europe has been the unwillingness to break with the past. I note with interest the heroic attempts in Great Britain and on the Continent, to continue to force this material into the Universal Decimal Classification. We found when the dust had settled after World War II here in America, that all of us who had been trying to operate on classified systems, hierarchical schemes, have had to give them up to greater or lesser degrees. The apparently radical departure from the Library of Congress decimal classification which Mr. Wall presented to you is really not radical at all. There are people present who have been operating very large systems for a good many years now without any trouble and with a very acceptable degree of economy moneywise. One of the features of this proposal which is very attractive, in my view, is the fact that you can start in a most modest fashion, but be building every step of the way on a firm foundation. I for one would like to express appreciation to Mr. Wall for an excellent analysis of a most interesting problem.

Mr. Topping: Someone, I believe it was the editor of an architectural magazine, said that the difference between a critique and an analysis was that a critique was for use with bold journalism, discussion of your contemporaries and rivals, whereas analysis was for use when there was danger of law suit, for all black tie affairs and for meetings of sweetness and light. Lets get going on some critique here. Mr. Gull, what do you think about what has been proposed for our system for documenting building research?

Mr. Gull: I am inclined to agree definitely with Mr. Frome and Dr. Sanford that the basic ideas expressed in Mr. Wall's paper are sound. So far as I am aware, there is no other system to which you can turn with any greater assurance of success than the one which he has presented. Research eventually might develop something, but this is likely to take many years. My concern, I'm afraid, has been in the larger picture which Mr. Topping ruled out at the beginning, but perhaps I might develop this within a restricted scope. You're already embarked in BRI upon the publication of an abstracting journal and this has an index at the end by which you can discover the topics which have been abstracted over the past several years. This index is a conventional, alphabetical index; Mr. Wall's index is also alphabetical but on the surface it appears to be less conventional. It seems to me that one of the things the membership of BRI needs to know is "Are the members of BRI willing to take ten or twenty minutes to learn how to use this different method to their own profit?" There are some mechanical techniques involved in the use of this index, and in published form this would present the unusual pattern of a split page. You turn the pages over until you match the two terms you want, then you begin to look for common numbers. This is known to be very effective in organizations such as Mr. Wall's where the staff making use of this kind of an index has already been trained in its use. If you go on with the publication of an ever larger abstracting bulletin, that is larger both in numbers of abstracts per year and in the number of subscribers and users, this is one of the considerations that you'll want to think about. Another will be the order—if you will—of the abstracts published. The implication of Mr. Wall's system is that they have a serially ordered arrangement. This does not necessarily mean that you have to arrange the abstracts serially in the publication—you can group them by products, building techniques, research developments and so forth, but the serially-numbered order should be maintained simply as an indexing method, to get you back from the index to the abstract itself. Finally, there is one feature of the type of system that Mr. Wall is proposing which I think augers well for the type of operation that you will be looking for in the future. If you develop a large abstracting service, approaching shall we say Chemical Abstracts in size, you will need a lot of extra equipment eventually and this type of system lends itself very well to mechanization. If you go in a different direction and offer an information service to which people write in for their answers, again this offers you the opportunity of a centralized location, and of going from a manual operation to a mechanical operation at some time in the future when load upon your service becomes great enough to warrant it.

Mr. Wall: There was one statement in my paper which may not have been picked up; that is when I stated that the indexer must be generally competent in the field of knowledge which he is indexing. This I glossed over and since nobody else has yet shot holes in my suggestions, allow me to take a pot shot. This may be a problem. It is in our culture apparently difficult to find a great many people satisfied to make a career of indexing technical literature—competent technical people who know the field which they would be indexing. At the same time, if you try to solve this problem by spreading the dirty work around, then you have the problem of giving these people the job as part-time, and of training more people than you would if it was done full-time by a certain number of people. There are also certain matters of consistency and variability which come up, so I think it is only fair to expound a little upon this matter of having competent indexing. It will have to be faced sooner or later by anybody who starts setting up a system. In fact, it has to be faced not only with the type of techniques and principles which I talked about this morning, but with any sort of retrieval system. The thing that makes it more critical in the method which I talked about this morning is that you are capable of what we call deep indexing with this method. It does not mean you have to index deeply, but you can if you find it desirable and when this happens the technical competence of the indexer has to be greater than if he were only going to put a book under the term architecture. Perhaps almost any clerk could decide that we are going to store this book on the architecture shelf, but when we start talking about deep indexing and we get down into the technical details, the further we go in that direction the more technical competence we need on the part of the indexers.

Mr. Miller: I think that no one is more of an advocate of this approach to indexing and retrieval than I am. I am a great believer in it. It has worked, and been very effective in most cases where it has been applied. I think there's a certain danger however, in beating the thing to death in respect to perfection, or seeking the perfection of it as a technique. Mr. Wall showed a series of curves this morning showing perfection as almost a function of dollar input. This isn't necessarily true, although the way he explained it, and the way he took, the point of maximum difference between effort and dollar was quite correct. But we can have reason to be concerned about what happens when we go beyond this point. Now, just to get really specific, let me pick up a small point and play with it for a minute; that is the point of role indicators. I believe in role indicators; I think that they are absolutely essential to a system where the subject matter is complex, where the possibility of false answers is high and where you get into problems as complex as both chemistry and patents. But I am not sure that at the outset the Building Research Institute collection is as complicated as this and needs elaborate role indicator or link structure in it. You don't get this for nothing; you pay a price. The more elaborate you get and the closer you approach this mythical point of perfection, the higher the cost, and you're beyond the point of diminishing returns. Now straight coordinate indexing, if the subject matter is not too complicated, can do quite a job. If you get a high percentage of false answers then obviously you've got to do something about it, but I don't think you should assume at the outset that you will. I don't think you ought to spend the money on a system so elaborate that you're not going to get a decent return on it. This is just a word of caution on this business of role indicators.

Mr. Whaley: I would like to go a little farther into this matter of depth of indexing and problems connected with the economics there. I won't start right out telling about our system, I promise you that, but I will say that the maintenance of a good index, and the building of one, are really two very well divided steps. The first one concerns the intellectual part of it; that is, the organization of your term file, your terminology, decisions as to depth, as to the need for it, and that means of course decisions as to whether a term should be divided into two or more terms. For example, in our work in chemistry, hydrogen looks like a good term until somebody finds you need to have the distinction between deuterium and hydrogen. Otherwise in a question on deuterium you are going to have 99% extraneous material dropped down. Somebody will want to pinpoint that under deuterium, so we have to introduce greater depth. We can also introduce greater depth by the introduction of roles and of links. Now this whole process is one that takes trained people, and when a person has made these decisions on terminology and built it up, then he actually comes to the job of indexing. This involves looking at a document, analyzing it, deciding the important points for future retrieval on the basis of your previous organization of terminology, adding to that organization if necessary and then filling out a work sheet or a form in which you identify the document and you also identify these terms with roles, etc. At the conclusion of this act the indexing itself is over. Here's where the manipulating process begins, and it is from this point on that many techniques are available. The statement has been made that there is only one theory and it can be applied in various ways. I think that Mr. Wall showed very well in his report that this theory at least can be subdivided into two immediate approaches whether you do this by the direct coding of your work sheet so that there's a single unit in your file for every document on which you've shown all the terms applying to it, or whether you invert that information on the work sheet so that your unit becomes a term or subject heading and under that you show all the documents pertaining to that term. Those are the two principal directions you can take at that point, and that's a separate problem—one of manipulation.

So we get back to the first one, which is the real big problem, the matter of organization and control of your terminology. It has to be maintained under very careful control, and that is where the question comes up of whether you should have permanent professionally trained indexers do it, or should dilute their activities so that you have a lot of people working only 10% of their time. I think its vastly superior to have a small group: it's much easier to keep that terminology under control. As an example, one of the branches of our corporation started out with the idea of every man there being one of the indexers, and just a small group of people trying to keep the terminology under control. That system did not work and after a few years it had to be abandoned. We can profit, I think, from some of that experience, and you can profit from it. My advice is to keep this group as small as possible.

The question has been brought up by Mr. Wall, "How can you get people to do this kind of work?" Well, I don't think it's any more monotonous or any more tedious than, say, analytical work. Nobody has any trouble hiring analytical people. I think that if we upgrade this type of work in our own organizations so that it's recognized as a branch of the profession in which you may be, then you shouldn't have too much trouble getting either men or women to do this work. What's the one thing that's the best incentive for all—money! If I have any exception to take to Mr. Wall's paper at all it would be on this point of dilution of the responsibility for maintaining and editing. I favor a smaller, carefully controlled group.

Mr. Heumann: I believe that the description that was given to you this morning is also a promising one for experimentation, but I have the feeling that one should be open to change in this field perhaps more than in many others. The computer part, for instance, has a life of only about 2 years now, and the machine that will eventually affect whatever you do is going to be affected by that kind of rapid replacement need. In addition to that, it seems to me that when you first set up a good information service you immediately find yourself in a little different situation than you were prior to setting it up. I think many people here have had the experience of beginning to provide good information to technical people and immediately finding that what these technical people wanted was a good deal more information, and more sophisticated information. These things have a kind of life of their own, and the situation five years hence, after you have satisfied what may be current uneasiness or unrest with regard to missing technical information may be quite different.

Mr. Frome: One thing that may be apparent, but I think should be stressed, is the accuracy of indexing. We have all assumed that once an indexer reads a document, and indexes it, if he is highly trained he doesn't miss very much. Well, experience has shown quite to the contrary, that in many cases an indexer does miss quite a bit of the material. The only insurance against this is to have it done possibly 2, 3, or 4 times depending on what you can afford. As a practical matter we have found that every document that's indexed, should be indexed at least twice by two separate people, because no system is any better than the accuracy of your indexing. One other point I'd like to make; I agree with Dr. Whaley that it is much better to have a corps of indexers whose sole job is that. This will give you closer control, and you can have two separate people index the same document. Then you can compare them and if there are inconsistencies you can find out which one is right and eliminate a lot of inaccuracies. Once a technical person finds your system has a lot of inaccuracies, or that he's missed things, you lose consumer acceptance.

Mr. Wall: I will be the first to admit that having a document indexed 2, 3, or 4 times will pick up additional points. I wouldn't like to say accuracy, I'd like to talk about this as depth of indexing or intensity. I am sure that another indexer, even using the same thesaurus as the first, will pick up additional points of interest in a document. I do feel however, that one reaches a point of diminishing returns very rapidly in this process. The first indexer may get 85 or 90% of it, the second one may get 85 or 90% of what the first one missed, and by the time you get to the third or fourth one, you've got hold of the short end of the lever economically.

Mr. Sanford: The position of the Patent Office is of course a peculiar one. It must have 100% retrieval if possible, because it is that one patent that might be in conflict with the patent application which must be found. In practice, using the coordinate indexing system, the percentage of information retrieval is so high that for practical purposes the level of sophistication, of education, of training of the indexer in my experience can be lower than for any other large and satisfactory system. I grant you that you can't get out of the system what you don't put in. However, the flexibility and the breadth of scope that are possible in this system make retrieval possible and habitual when in ordinary terms of carelessness of the indexing, the operator of the system really feels he has no business to be able to recover it. I run a government organization in which I have to use all kinds of people doing indexing, and I know from sad experience that I can operate this system with people of much less skill and much less training, and operate it more satisfactorily than any other class of system which we have tried. The system has another marked advantage, and that is its speed. It is possible to get through a great deal of work in a very short time. I think our emphasis ought to be on the sophistication of the retrieval operation rather than on the sophistication of the storage operation. It's a truism in this game that you're going to retrieve a very small percentage, ever, of the total that you store. Therefore, if your storage operation is expensive your costs go way, way, up. If your storage operation is cheap and you spend your time, energy and money on the retrieval operation to make that more sophisticated, you get much more for your money.

Mr. Heilprin: To comment further on the point that has just been made, we all recognize that there are two kinds of access to an information system. One of them is the intellectual access to an idea, the other is the physical access. Very frequently one gets up a good intellectual scheme, but finds that the physical system that one thinks about as a model to embody it has some flaws which do not show at first, but later make it a rather difficult system to operate. So I'd like to reinforce what Dr. Sanford has just said. He used the words "minimum control." The words I would use would be "do not inadvertently introduce constraint of which you aren't aware." For instance, you make the simplest kind of a decision; say to use an abstract, or a 3" x 5" card instead of a 4" x 7" card. Later you may find that because of the size of this thing there are whole classes of machines that you could use or you couldn't use that would, on the one hand, cost you a great deal to convert to, and in the other case would not, etc. As I mentioned there are two kinds of access in the process of retrieval, the intellectual and the physical. The physical thing is the thing that we all must cope with in the actual work of getting information out of the system, and there, the physical form of the record depends upon the techniques which are available at the time. As we all know the most usual form of storing records is to store the whole document, the abstract, perhaps line titles. Now supposing that we said we were going to select one of these at a given time, a few years from now, we may find that it is possible to transmit images by facimile or TV or other means at a rate that is quite sufficient so that you could have a small file some place and would not need a condensed abstract; you could transmit the whole document if you wanted. If you made your choice on the basis of the fact that the abstract was physically small, later you might find that it would have been just as easy to store 50 or 100 pages in the same space. On the other hand, if you made your decision on the basis that the user, scientist or engineer, is not going to want to scan more than the amount that is in an abstract, then you may find that your original decision was correct, and that the physical form is less quick. In the science called "Operations Research" we are always looking for what is called optimalization. Now optimalization means many things, but essentially what it needs is a measure of effectiveness. What you use as a measure of effectiveness in selecting a system for storage and retrieval may differ as time goes on. I gather the membership of this group at the present time is partly professional people and partly people in the industrial and contractual line of work. If they consult these documents they do not do so in the same way that a scholar would. They want some information quickly. It is possible that, as a result of using this system, the entire group of building research people may change into a much more, shall we say, professional scientific group. They may become more aware of the scientific elements in their work and, as was mentioned, the depth of indexing required will be much greater. So, I simply say that what you optimalize on at one time may change. Of course the more common measures of effectiveness in information system retrieval are the access time, the storage cost, the security against destruction, security meaning secrecy, and also the dissemination power of this system to get information distributed. All these principal measures of effectiveness, and others perhaps more personal or more necessary to you, should be considered from the standpoint—are they permanent, are they going to wear, are they going to change.

Mr. Topping: As has been mentioned, BRI has a small abstracting service which is indexed only by subject. We are not certain just where we are going to go from here, but we did question our membership on whether they found this abstract service useful and asked for their criticisms of it. The answer came back almost universally that they found it useful, and almost 100% found it hard to find things without a lot of search, indicating that what was needed was an indexing system. We recognize that an indexing or retrieval system doesn't have to be very sophisticated to take care of 400 documents a year, but on the other hand we don't intend to limit our coverage to such a small number as this. We may continue to send abstracts to our members as we do now of articles that are not covered in other indexing services, but it would also be entirely logical to coordinate and blend in with this a method of retrieving information from other abstracting services such as ACI. Glenn Beyer listed a great many disciplines which would obviously have pertinent literature. This is one of the things we are going to have to reconcile, so that in discussing the manner of handling our retrieval system we are still in a pretty fluid state. One thing that interested me was the statement by Dr. Sanford that the number of retrievals is less than the number of items that you put in. Would this be characteristic of only certain types of information? We think that if the number of retrievals is small compared to the number of entries, we ought not to do the job. If the retrievals are not high we cannot afford to do it.

Mr. Sanford: Unfortunately the very nature of research is bumping into the unexpected in the exploitation of the unknown. The percentage of retrieval from the total store is bound to be low because you just can't guess that well. This is true of any information system, including the Library of Congress where there are a good many millions of books of which many millions have probably never been off the shelf since they were put there. But that doesn't mean that they shouldn't be there, because some day they may be needed. This whole concept of efficiency in terms of a shoe button factory just doesn't apply to the field of research, as you well know.

Richard Mann, Prevention of Deterioration Center, NAS: Should prepared abstracts be meaty and technical or should they merely describe what can be found in the original document? Should any abstract be so well written that reference to the original document is made unnecessary?

Mr. Wall: This, again, is one of the things I call a tactical situation. Our work within a department of a single company makes us feel that for our purposes the abstracts should be indicative; that is, they should state what may be learned from the document, but not attempt to be fully informative. To me it seems that we are asking for the epitome of literary and technical skill in asking that an abstract of a very limited number of words be developed which summarizes a document when some author has found a whole article or paper necessary to get his point across then we ask for an abstract to be prepared in a few words which says everything that he said in his paper. This would be an informative abstract. We have felt that for our purposes we should have indicative abstracts and we do not wish the people in our organization to work from the abstracts alone. We merely want to use the abstracts as a method for letting them screen further the documents to which they are referred so that they can tell which ones to look at first, which ones to look at later and which ones they perhaps already know about. On the other hand, our organization makes available to every customer the documents themselves when they want them. This is not the situation in many other places. I believe that Chemical Abstracts attempts to be informative on the basis that not everybody who employs Chemical Abstracts has available, or can get easily, the documents from which the abstracts were made. Therefore, they must attempt to be informative because people might have to work from the abstracts alone. It seems to me that the decision on what kind of abstracts to use must be made on this basis: Are your clients going to be able to put their hands easily on the original documents, on the complete documents, or are a sizeable number of them likely to work from the abstracts alone? If they can put their hands on the complete documents then I think they ought to have indicative abstracts; if they can't, then you must have informative abstracts.

Mr. Miller: I think that any money spent on indicative abstracts is wasted. An abstract has two functions; the indicative abstract is one which is an additional aid to search or screening; the other type, the informative abstract, is a substitute for the original document. The Chemical Abstract type is in this category. I don't know how many items they abstract, probably 20 or 30 thousand different types of things, but none of us can look at them all. To have a substitute for the original is certainly an acceptable, in fact a necessary thing. The reason I seriously believe that the indicative abstract is a waste of money, is that if you have an adequate indexing organization and you have an adequate retrieval structure, designed so that you can be as specific or as general as you like, then this apparatus should take you directly to the document. Why interpose at some expense another screening device between the index and the document?

Mr. Whaley: Informative abstracts have the danger of leaning towards the point of view of the abstracter. This is not bad if the clientele has the same point of view as the abstracter. For example, Chemical Abstracts makes a great point of the fact that all new compounds shown in an article will be shown in the abstract, so anybody looking for new compounds can be certain that that abstract is just as informative as the original article with respect to that point of view. However, Biochemical Abstracts, not particularly interested in that point of view, will abstract all the biochemical information and the same article will have two different informative abstracts, or may still have a third one in Nuclear Science Abstracts. The point is, if you're going to try to be informative you have to limit yourself. No abstract can give information on everything that's in the original article; by definition it can't, because if it could then it would be the original article, and somebody's done a lot of padding in that original article, if there's an informative abstract that can do the same job. However, the indicative abstract takes a different approach. It helps to bolster a shallow index; that is, if you have an index that doesn't go very deep and you are going to get a lot of extraneous material as a result of it. It would be helpful if you could look at an indicative abstract, and go down the list of it, and see from that which of these articles are really worth getting out and are going to bear directly upon the situation, rather than getting a hundred of them out of which only 10 are necessary. Mr. Miller brought up the point that if you really have a good deep index, you don't have to do that. He previously said that it's difficult to have a deep one, because it costs money, and it does. So, you can either spend your money on a deep index and not require an indicative abstract, or have a shallow index and need an indicative abstract, which again is expensive.

Mr. Sanford: The indicative abstract is normally produced for another purpose. It's for the "newspaper" purpose, to get the word around fast, inviting the attention of the people who should look at the full text. This is its normal function as part of a retrieval system, not as something that is a substitute for the indexing.

Mr. Mann: Strictly from the standpoint of furnishing complete data to the requester of information, do you favor one large centralized information center for the entire U.S. or a large number of small information centers dealing with specific areas?

Mr. Miller: This is a very large question and there are a lot of things that would impinge on the correct answer, including the cost, the economics, and what you get out of it. Maybe because I come from a very small state, Vermont, I sort of abhor bigness. I don't really like large information centers. My feeling is that the specialty shops are a lot better shopping grounds for what you really want from them than the large department stores.

Mr. Gull: My belief is that the question posed here cannot be answered without a great deal of study. It is a political question. We have seen a

Russian model of a centralized one reported only this week in Science as tending toward the decentralized situation, compared to our own generally expressed political views in favor of decentralization. You simply cannot give a clear-cut answer to this without studying the situation out. Even if a commission were set up to study this now, and worked for five years and came up with an answer, you might find that the technology at the end of the five years would enable you to do quite different things than you can do now. This is a point that I think Dr. Heumann has already emphasized.

Mr. Miller: May I make one more point on this. I guess I'm inclined toward several special information center. But at that point another responsibility must be taken up by the operators of those centers, and that is to arrange a really effective interchange among them so that all of the several disciplines, or any one discipline using one, can get the benefit of them all. I think this would be the most satisfactory form.

Mr. Topping: I believe it is the feeling of the BRI Research Committee that we should not do anything just because we are the Building Research Institute. If anyone else can do it better or has the resources we would by all means be delighted to see them do it. We have plenty of work to do without taking on more. But this job needs doing, and if a big central agency would do it, and it would work for us, I think we would be less inclined to set up a separate one ourselves.

Mr. Wall: Several people on the panel have been concerned about constraints, or limitations, or fixed positions in these storage and retrieval systems. I think we should clarify this jargon a bit for our audience. I'd like to give examples of the two major constraints which I talked about this morning. The first of them is that of the thesaurus, which in effect is a control of the vocabulary. This is one constraint which I am unwilling on principle to move away from without extremely strong proof that it is not necessary, in order to afford effective communication among the originator, indexer and user of information. The other major constraint is that of relationships, role indicators, subdivisions of the documents, etc. We have gone into these because it has become apparent that they are not really expensive, and we can always dispense with them very easily and use something somewhat simpler. But if we don't use them now, and we ever decide we do want to, it is very difficult to start using them and to mesh your previous index with an index which does use role indicators and document subdivisions. So, we've thought it the better part of valor to use them at this stage of the game. On the matter of indicative abstracts costing money and not being necessary, I do feel that the index which will lead you directly to the document without any error whatsoever is perhaps a will-o-the-wisp, that it will never be attained, and that there will always be a desire to have some intermediate screening operation. When you index deeply, with relationships such as role indicators included, you can create an indicative abstract automatically from the indexing. It will be a stylized sort of pidgin English, but the use of the role indicators, etc., will enable you to create an indicative abstract mechanically from the indexing itself.

Mr. Topping: I want to get in three or four questions here: Is it feasible for an organization the size of BRI to do this, or can it be done by a large organization, something of the size that Mr. Frome has in the Patent Office? Is it within our scope to start on something like this and, presumably if we are successful, expand it, or is it foolish to tackle a collection of disiplines as large as this with a small organization such as we have?

Mr. Heilprin: I'd like to answer that the other way around. Unless you who are interested in these subjects, from your point of view, start it, it probably won't get done. So any size with which you start would be the best size.

Mr. Heumann: I certainly think that every abstracting periodical at some time started with one abstract, the first one they ever did. Chemical Abstracts started with a journal list of about 100 titles; it now does about 8,000 regularly. While you may feel your own problems are unique, and that your coverage and scope are going to be so different from others', I would be surprised if you couldn't find a lot of valuable experience to be picked up from other operating abstracting services, even though their subject matter may seem to be quite different from your own. All these people went through a lot of headaches at one time or another on this matter of growth and determining the boundaries of their subject. You can get a lot of valuable free advice by going to people who have already been through this.

Mr. Frome: One thing that impressed me when I was listening to the first speaker, and also to the comments of the chairman, is "our problem is different, our problem is immense, our problem crosses many disciplines." If you had attended as many documentation meetings as I have, you would know that everybody who has a problem comes up with the same story. In fact, the stories are all true. The subject matter may be very different, but the problems are basically the same; the exponential bulk of literature, of specialization, specialized jargon, what to do about it. The important factor in deciding whether BRI can do it themselves or not is motivation. They are the ones that are interested in doing it: there exists a problem of communication. You can hire a documentation specialist, but it's a very great problem to communicate to him what you really want. Many times you don't know. At least in the beginning the people who want the service should be the ones to do it. After it is well defined, you have stated very clearly what you want and you know what you want, then you may turn it over to someone else, but certainly at the beginning the people who need the service should be the ones to do it, with the help of professional documentalists.

Mr. Miller: Yes, I would like to second what Mr. Frome said. I came out of the aeronautical discipline where we really thought we covered the universe. In the design of an aeroplane we used all the same terms that Dr. Beyer did this morning. We were concerned about acoustics, sound, lighting, garbage disposal, psychology, human factors and motivation.

Mr. Topping: Mr. Miller has brought up the point that the aeroplane industry uses the same phrases, the same disciplines, and he has found others who do. Does this mean, then, that if we documented the building industry we would be duplicating what other people are doing, or that it would overlap by 60 or 80% and, if so, should we not do it, but put it in somebody else's hands?

Mr. Miller: This is why I made my plea belatedly a moment ago, for an interchange among the different documentation centers. I think it's amazing how much one center can contribute to the needs of another, and they ought to get together more often. They will find that they have a tremendous community of interests. If you are dealing with scientific subjects, you are dealing with all sciences. It isn't necessarily duplication, probably isn't duplication of your approach to organizations, but I think that there is a common interest here that could be explored, and should be explored the minute you start a documentation center of any size.

Mr. Topping: You're making a distinction here between cooperation between centers andthe condensation of several interests in one documentation center. You're saying it would be better to have several documentation centers and for them to cooperate rather than for us to join the aeroplane industry, for instance.

Mr. Miller: I think that if you joined the aeroplane industry in this documentation effort you might find that you are talking about the same thing in a different language occasionally. You might find that they are dealing in a lot of dynamics that you're not dealing with. What you need in most of these centers is a switching point which enables one to communicate in the terms of the other.

Mr. Wall: We of the same disciplines could also be found in a number of other industries, other than the aircraft industry and building research. In fact, I think you will find little duplication between the aircraft industry and building research on the actual items or documents that are chosen for indexing. I think people know when a document is pertinent to the aircraft industry and they know when it is pertinent to the building industry. I don't think you will find much duplication in that respect. On fundamental research, strength of materials and things like that, there may be some things of interest to each, but by and large I would be very surprised to find a great overlap between the items in the collection of these two separate centers.

Mr. Topping: How do you decide when you have to mechanize? Is it a question of economics—a question of when the cost of operating the thing manually begins to be bigger than the cost of operating it mechanically? Or what is it?

Mr. Heilprin: Well it's partially a matter of economics, but there is another practical consideration. When the collection gets to a certain size, so that the time to get a piece out begins to be a little bit long and the members of the organization begin to grumble a bit, you realize that the access time has gone beyond practical limits and you have to do something. At that point you may not yet have to mechanize; you may find that examination of the system will introduce means of increasing the speed, but in general there is a reasonably well defined point that you yourself will determine when it will begin to be heavy and unwieldy. That will be your signal—the money, of course, will be very obvious.

Mr. Frome: Although economics is very important, there are instances where time of access is of extreme importance: if you don't have the information in the next five minutes or half hour you don't need it at all, it's too late. So although cost is a prime factor many times, the period of time becomes extremely important.

Mr. Wall: May I make one comment on mechanization. When one reaches the point where something has to be done to speed things up or to cut the cost, this is not a jump-off-the-skyscraper situation. You can mechanize by stages with a properly designed system. Some very simple machines exist for the early stages of mechanization—low-cost machines—and you can go on up, as Mr. Miller pointed out this morning, to the extremely expensive computers if you have to.

Mr. Gull: There will also be, at times, a situation in which a machine will enable you to do something which you can't attempt in a manual system. You may have to start with an electronic or mechanical system to get into operation because your requirements are so high to begin with. I am of the opinion that you in BRI are at a situation where you can start with a small manual or publishing level and work towards a larger one, but this isn't true for everybody.

PREVIOUSLY PUBLISHED BUILDING RESEARCH INSTITUTE CONFERENCE PROCEEDINGS

PLASTICS IN BUILDING, 1955, 150 pages, illustrated, Publication No. 337, $5.00.

METAL CURTAIN WALLS, 1955, 190 pages, illustrated, Publication No. 378, $4.00.

FLOOR-CEILINGS AND SERVICE SYSTEMS IN MULTI-STORY BUILDINGS, 1956, 141 pages, illustrated, Publication No. 441, $4.00.

MODERN MASONRY, NATURAL STONE AND CLAY PRODUCTS, 1956, 163 pages, illustrated, Publication No. 385, $4.50.

WINDOWS AND GLASS IN THE EXTERIOR OF BUILDINGS, 1957, 176 pages, illustrated, Publication No. 478, $5.00.

ADHESIVES AND SEALANTS IN BUILDING, 1958, 160 pages, illustrated, Publication No. 577, $5.00.

INSTALLATION AND MAINTENANCE OF RESILIENT SMOOTH-SURFACE FLOORING, 1959, 146 pages, illustrated, Publication No. 597, $5.00.

FIELD APPLIED PAINTS AND COATINGS, 1959, 140 pages, illustrated, Publication No. 653, $5.00.

NOISE CONTROL IN BUILDINGS, 1959, 136 pages, illustrated, Publication No. 706, $5.00.

SEALANTS FOR CURTAIN WALLS, 1959, 82 pages, illustrated, Publication No. 715, $3.00.

BUILDING ILLUMINATION: THE EFFECT OF NEW LIGHTING LEVELS, 1959, 106 pages, illustrated, Publication No. 744, $5.00.

BUILDING RESEARCH, INTERNATIONAL, 1960, 41 pages, Mimeo, $1.50.

NEW METHODS OF HEATING BUILDINGS, 1960, 142 pages, illustrated, Publication No. 760, $5.00.

These publications are available on order from Publications Office, National Academy of Sciences-National Research Council, 2101 Constitution Avenue, N.W., Washington 25, D.C.
A full list of BRI publications is available on request from the Building Research Institute at the above address.

BUILDING RESEARCH INSTITUTE

The Building Research Institute is a unit of the Division of Engineering and Industrial Research of the National Academy of Sciences-National Research Council. BRI was organized in 1952 to meet the need of the construction industry for an organization which could focus the attention of the entire industry on building research and technology. It also acts as an information center and maintains liaison with building research agencies in other countries throughout the world.

The members of BRI are people interested in advancement of the science of building. Among those listed as BRI members are: architects, engineers, contractors, home builders, building owners, manufacturers of building products and materials, distributors, technical and professional societies, trade associations, research laboratories, financial, real estate and insurance firms, trade and consumer publications, professional consultants and technical experts from colleges, universities and government agencies in this country and abroad. Memberships are open to companies, associations, societies and individuals.

MEETINGS

Operating on the principle that the personal exchange of experience and ideas is the basis of the growth of a science, BRI conducts:

1) Research correlation conferences on specific design problems and the cross-industry application of building products (Open to the public)

2) Workshop, round-table and study groups on specific subjects (Open to BRI members with invited guests)

Research correlation conferences take the form of multi-subject meetings and are held twice a year, spring and fall. Programs on various subjects of interest to the building industry and its related professions of architecture and engineering are presented in half-day, full-day, two-day or three-day sessions, depending on the field to be covered and the amount of time necessary.

PUBLICATIONS

The Building Research Institute publishes and distributes to members the proceedings of its conferences, technical meetings and study groups. Building Science News, the Institute newsletter, reports monthly on Institute activities, as well as on building research news of general interest. Building Science Directory, founded in 1956, provides a comprehensive guide to sources of information on research and technical developments in the industry. Supplements to the Directory are issued quarterly with an annual index. BRI Abstracts of Building Science Publications are published quarterly. All of these services are provided to BRI members without charge. Non-members may purchase copies of published proceedings of public conferences and regular issues of the Building Science Directory at nominal cost.