CHAPTER

Advanced Web Searching

5

INFORMATION IN THIS CHAPTER

- Search Engines
- Conventional Search Engines
- Advanced Search Operators of various Search Engines
- Examples and Usage

INTRODUCTION

In the last chapter we dealt with some special platforms which allowed us to perform domain-specific searches; now let's go into the depths of conventional search engines which we use on daily basis and check out how we can utilize them more efficiently. In this chapter, basically, we will understand the working and advanced search features of some of the well-known search engines and see what all functionalities and filters they provide to serve us better.

So we already have a basic idea about what search engine is, how it crawls over the web to collect information, which are further indexed to provide us with search results. Let's revise it once and understand it in more depth.

Web pages as we see them are not actually what they look like. Web pages basically contain HyperText Markup Language (HTML) code and most of the times some JavaScript and other scripting languages. So HTML is basically a markup language and uses tags to structure the information, for example the tag <h1></h1> is used to create a heading. When we receive this HTML code from the server, our browsers interpret this code and display us the web page in its rendered form. To check the client-side source code of a web page, simply press Ctrl+U in the browser with a web page open.

Once the web crawler of a search engine reaches a web page, it goes through its HTML code. Now most of the times these pages also contain links to other pages, which are used by the crawlers to move further in their quest to collect data. The content crawled by the web crawler is then stored and indexed by search engine based on variety of factors. The pages are ranked based upon their structure (as defined in HTML), the keywords used, interlinking of the pages, media present on the page, and many other details. Once a page has been crawled and indexed it is ready to be presented to the user of the search engine depending upon the query.

Hacking Web Intelligence. http://dx.doi.org/10.1016/B978-0-12-801867-5.00005-7

Once a page has been crawled, the job of the crawler does not finish for that page. The crawler is scheduled to perform the complete process again after a specific time as the content of the page might change. So this process keeps on going and as new pages are linked they are also crawled and indexed.

Search engine is a huge industry in itself which helps us in our web exploration, but there is another industry which depends directly on search engines and that is search engine optimization (SEO). SEO is basically about increasing the rank of a website/web page or in other words to bring it up to the starting result pages of a search engine. The motivation behind this is that it will increase the visibility of that page/site and hence will get more traffic which can be helpful from a commercial or personal point of view.

Now we have a good understanding of the search engines and how they operate, let's move ahead and see how we can better use some of the conventional search engines.

GOOGLE

Google is one of the most widely used search engines and is the starting point for web exploration for most of us. Initially Google search was accessible through very simple interface and provided limited information. Apart from the search box there were some special search links, links about the company, and a subscription box where we could enter our email to get updates. There were no ads, no different language options, no login, etc.

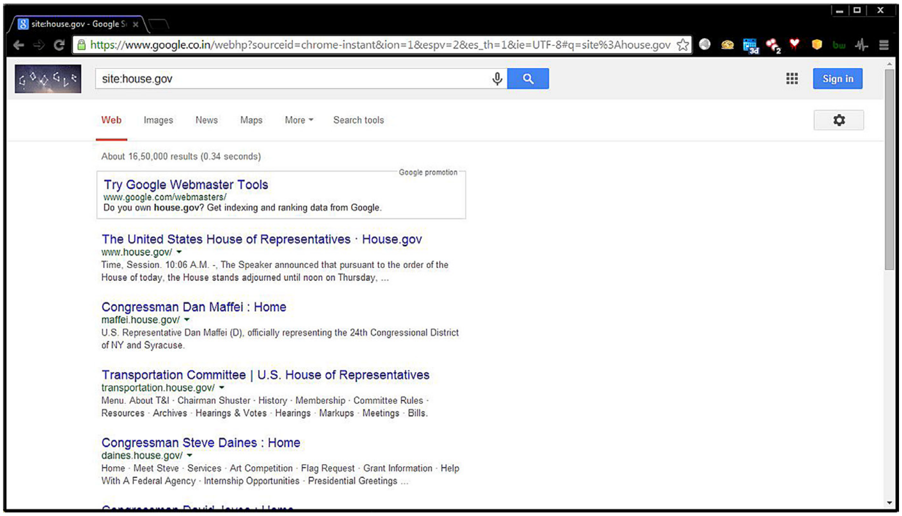
It's not only the look and feel of the interface that has changed over the years but also the functionalities. It has evolved from providing simple web links to the pages containing relevant information to a whole bunch of related tools which not only allow us to search different media types and categories but also narrow down these results using various filters. Today there are various categories of search results such as images, news, maps, videos, and much more. These plethora of functionalities provided by Google today has certainly made our lives much easier and made the act of finding information on the web a piece of cake. Still sometimes we face difficulty in finding the exact information we are looking for and the main reason behind it is not the lack of information but to the contrary the abundance of it.

Let's move on to see how we perform Google search and how to improve it. So whenever we need to search something in Google we simply think about some of the keywords associated with it and type them into the search bar and hit Enter. Based upon the indexing Google simply provides us with the associated resources. Now if we want to get better results or filter the existing results based upon various factors, we need to use Google advanced search operators. Let's have a look at these operators and their usage.

site:

It fetches results only for the site provided. It is very useful when to limit our search to some specific domain. It can be used with another keyword and Google will bring back related pages from the site specified. For an information security perspective it is very useful to find out different sub domains related to a particular domain.

Examples: site:gov, site:house.gov


site:house.gov - Google S
https://www.google.co.in/webhp?sourceid=chrome-instant&ion=1&espv=2&es_th=1&ie=UTF-8#q=site%3Ahouse.gov
site:house.gov
Sign in
Web
Images
News
Maps
More
Search tools
Google promotion

About 16,50,000 results (0.34 seconds)

Try Google Webmaster Tools
www.google.com/webmasters/
Do you own **house.gov**? Get indexing and ranking data from Google.

The United States House of Representatives · House.gov
www.house.gov/
Time, Session. 10:06 A.M. -, The Speaker announced that pursuant to the order of the House of today, the House stands adjourned until noon on Thursday, ...

Congressman Dan Maffei : Home
maffei.house.gov/
U.S. Representative Dan Maffei (D), officially representing the 24th Congressional District of NY and Syracuse.

Transportation Committee | U.S. House of Representatives
transportation.house.gov/
Menu. About T&I · Chairman Shuster · History · Membership · Committee Rules · Resources · Archives · Hearings & Votes · Hearings · Markups · Meetings · Bills.

Congressman Steve Daines : Home
daines.house.gov/
Home · Meet Steve · Services · Art Competition · Flag Request · Grant Information · Help With A Federal Agency · Internship Opportunities · Presidential Greetings ...

FIGURE 5.1

Google "site" operator usage.

inurl:

This operator allows looking for keywords in the uniform resource locator (URL) of the site. It is useful to find out pages which follow a usual keyword for specific pages, such as contact us. Generally, as the URL contains some keywords associated with the body contents, it will help us to find out the equivalent page for the keyword we are searching for.

Example: inurl:hack

allinurl:

Similar to "inurl" this operator allows looking for multiple keywords in the URL. So we can search for multiple keywords in the URL of a page. This also enhances the chances of getting quality content of what we are looking for.

Example: allinurl:hack security

intext:

This operator makes sure that the keyword specified is present in the text of the page. Sometimes just for the sake of SEO, we can find some pages only contain keywords to enhance the page rank but not the associated content. In that case we can use this query parameter to get the appropriate content from a page for the keyword we are looking for.

Example: intext:hack

allintext:

Similar to the "intext" this operator allows to lookup for multiple keywords in the text. As we discussed earlier the feature of searching for multiple keywords always enhances the content quality in the result page.

Example: allintext:data marketing

intitle:

It allows us to restrict the results by the keywords present in the title of the pages (title tag: <title>XYZ</title>). It can be helpful to identify pages which follow a convention for the title of the pages such as directory listing by the keywords "index of" and most of the sites provide the keywords in the title for improving the page rank. So this query parameter always helps to search for a particular keyword.

Example: intitle:blueocean

allintitle:

This is the multiple keyword counterpart of "intitle" operator.

Example: allintitle:blueocean market

filetype:

This operator is used to find out files of a specific kind. It supports multiple file types such as pdf, swf, kml, doc, svg, txt, etc. This operator comes handy when we are only looking for specific type of files on a specific domain.

Example: filetype:pdf, site:xyz.com, filetype:doc

ext:

The operator ext simply stands for extension and it works similar to the filetype operator.

Example: ext:pdf

define:

This operator is used to find out the meaning of the keyword supplied. Google returns dictionary meaning and synonyms for the keyword.

Example: define:data

AROUND

This operator is helpful when we are looking for the results which contain two different keywords, but in close association. It allows us to restrict the number of words as the maximum distance between two different keywords in the search results.

Example: A AROUND(6) Z

AND

A simple Boolean operator which makes sure keywords on both the side are present in the search results.

Example: data AND market

OR

Another Boolean operator which provides search results that contain either of the keyword present on both the sides of the operator.

Example: data OR intelligence

NOT

Yet another Boolean operator which excludes the search results that contain the keyword followed by it.

Example: lotus NOT flower

“”

This operator is useful when we need to search for the results which contain the provided keyword in the exact sequence. For example we can search pages which contain quotes or some lyrics.

Example: “time is precious”

-

This operator excludes the search results which contain the keyword followed by it (no space).

Example: lotus -flower

*

This wildcard operator is used as a generic placeholder for the unknown term. We can use this to get quotes which we partially remember or to check variants of one.

Example: “* is precious”

..

This special operator is used to provide a number range. It is quite useful to enforce a price range, time range (date), etc.

Example: japan volcano 1990..2000

info:

The info operator provides information what Google has on a specific domain. Links to different types of information are present in the results, such as cache, similar websites, etc.

Example: info:elsevier.com

related:

This operator is used to find out other web pages similar to the provided domain. It is very helpful when we are looking for websites which provide similar services to a website or to find the competitors of it.

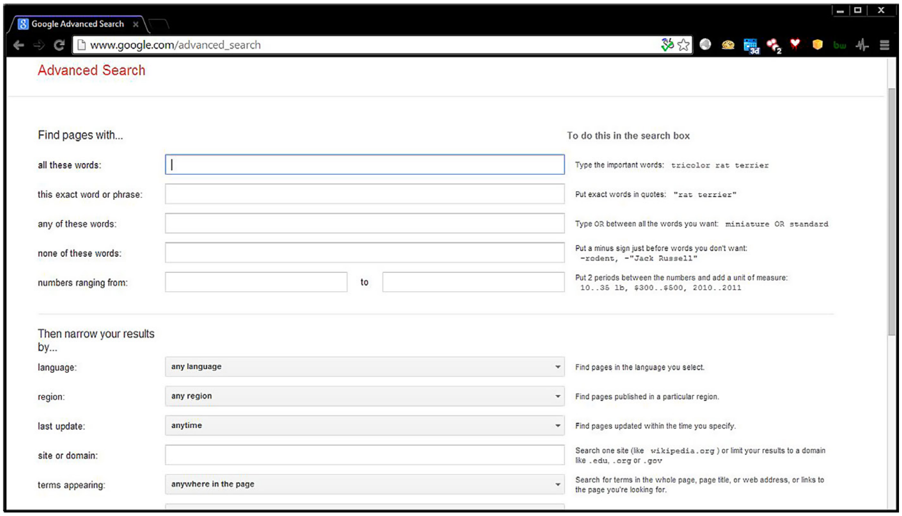
Example: related:elsevier.com

cache:

This operator redirects to the latest cache of the page that Google has crawled. In case we don't get a result for a website which was accessible earlier, this is a good option to try.

Example: cache:elsevier.com

Advanced Google search can also be performed using the page http://www.google.com/advanced_search, which allows us to perform restricted search without using the operators mentioned above.



Google Advanced Search

www.google.com/advanced_search

Advanced Search

Find pages with...		To do this in the search box
all these words:		Type the important words: `tricolor rat terrier`
this exact word or phrase:		Put exact words in quotes: `"rat terrier"`
any of these words:		Type OR between all the words you want: `miniature OR standard`
none of these words:		Put a minus sign just before words you don't want: `-rodent, -"Jack Russell"`
numbers ranging from:	to	Put 2 periods between the numbers and add a unit of measure: `10..35 lb, $300..$500, 2010..2011`

Then narrow your results by...		
language:	any language	Find pages in the language you select.
region:	any region	Find pages published in a particular region.
last update:	anytime	Find pages updated within the time you specify.
site or domain:		Search one site (like `wikipedia.org`) or limit your results to a domain like `.edu`, `.org` or `.gov`
terms appearing:	anywhere in the page	Search for terms in the whole page, page title, or web address, or links to the page you're looking for.

FIGURE 5.2

Google advanced search page.

Apart from the operators Google also provide some operations which allow us to check information about current events and also perform some other useful things. Some examples are:

time

Simply entering this keyword displays the current time of the location we are residing in. We can also use name of region to get its current time.

Example: time france

weather

This keyword shows the current weather condition of our current location. Similar to "time" keyword we can also use it to get the weather conditions of a different region.

Example: weather sweden

Calculator

Google also solves mathematical equations and also provides a calculator.

Example: 39*(9823-312)+44/3

Convertor

Google can be used to perform conversions for different types of units like measurement units, currency, time, etc.

Example: 6 feet in meters

This is not all, sometimes Google also shows relevant information related to global events as and when they happen; for example, FIFA World Cup.

Apart from searching the web, in general, Google also allows us to search specific categories such as images, news, videos, etc. All these categories, including web have some common and some specific search filters of their own. These options can simply be accessed by clicking on the "Search tools" tab just below the search bar. We can find options which allow us to restrict the results based upon the country, time of publish for web; for images there are options like the color of image, its type, usage rights, etc. and similarly other relevant filters for different categories. These options can be very helpful in finding the required information of a category as they are designed according to that specific category. For example if we are looking for an old photograph of something it is a good idea to see only the results which are black and white.

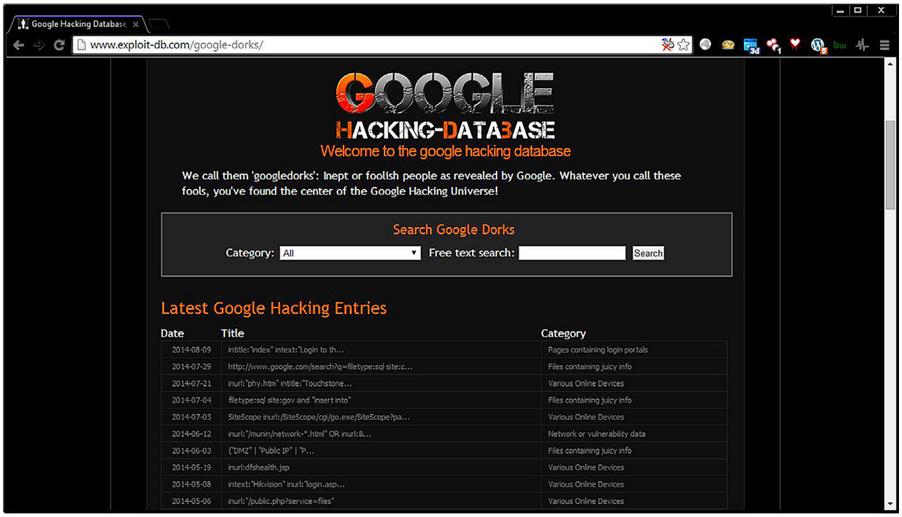
The operators we discussed are certainly very useful for anyone who needs to find out some information on the web, but the InfoSec community has certainly taken it to next level. These simple and innocent operators we just discussed are widely used in the cyber security industry to find and demonstrate how without even touching the target system, critical and compromising information can be retrieved. This technique of using Google search engine operators to find such information is termed as "Google Hacking."

When it comes to "Google Hacking" one name that jumps out in mind is Johnny Long. Johnny was an early adopter and pioneer in the field of creating such Google queries which could provide sensitive information related to the target. These queries are widely popular by the name Google Dorks.

Let's understand how this technique works. We saw a number of operators which can narrow down search results to a specific domain, filetype, title value, etc. Now in Google Hacking our motive is to find sensitive information related to the target; for this people have come up with various different signatures for different files and pages which are known to contain such information. For example, let's just say we know the name of a sensitive directory which should not be directly accessible to any user publicly, but remains public by default after the installation of the related application. So now if we want to find out the sites which have not changed the accessibility for this directory, we can simply use the query "inurl:/sensitive_directory_name/" and we will get a bunch of websites which haven't changed the setting. Now if we want to further narrow it down for a specific website, we can combine the query with the operator "site," as "site:targetdomain.com inurl://sensitive_directory_name/." Similarly we can find out sensitive files that are existing on a website by using the operators "site" and "filetype" in collaboration.

Let's take another example of Google Hacking which can help us to discover high severity vulnerability in a website. Many developers use flash to make websites more interactive and visually appealing. Small web format (SWF) is a flash file format used to create such multimedia. Now there are many SWF players known to be vulnerable to cross-site scripting (XSS), which could lead to an account compromise. Now if we want to find out if the target domain is vulnerable to such attack, then we can simply put in the query "site:targetdomain.com filetype:swf SWFPlayer_signature_keyword" and test the resulting pages using publicly available payloads to verify. There are huge number of signatures to find out various types of pages such as sensitive directories, web server identification, files containing username/password, admin login pages, and much more.

The Google Hacking Database created by Johnny Long can be found at http://www.hackersforcharity.org/ghdb/ though it is not updated, yet it is a great place to understand and learn how we can use Google to find out sensitive information. A regularly updated version can be found at http://www.exploit-db.com/google-dorks/.


Google Hacking Database
www.exploit-db.com/google-dorks/
GOOGLE
HACKING-DATABASE
Welcome to the google hacking database

We call them 'googledorks': Inept or foolish people as revealed by Google. Whatever you call these fools, you've found the center of the Google Hacking Universe!

Search Google Dorks

Category: All ▾ Free text search: Search

Latest Google Hacking Entries

Date	Title	Category
2014-08-09	intitle:"index" intext:"Login to th...	Pages containing login portals
2014-07-29	http://www.google.com/search?q=filetype:sql site:c...	Files containing juicy info
2014-07-21	inurl:"phy.htm" intitle:"Touchstone...	Various Online Devices
2014-07-04	filetype:sql site:gov and "insert into"	Files containing juicy info
2014-07-03	SiteScope inurl:/SiteScope/cgi/go.exe/SiteScope?pa...	Various Online Devices
2014-06-12	inurl:"/munin/network-*.html" OR inurl:&...	Network or vulnerability data
2014-06-03	("DMZ" | "Public IP" | "P...	Files containing juicy info
2014-05-19	inurl:dfshealth.jsp	Various Online Devices
2014-05-08	intext:"Hikvision" inurl:"login.asp...	Various Online Devices
2014-05-06	inurl:"/public.php?service=files"	Various Online Devices

FIGURE 5.3

Google hacking database- www.exploit-db.com/google-dorks/.

BING

Microsoft has been providing search engine solutions from a long time and they have been known with different names. Bing is latest and most feature-rich search engine in this series. Unlike its predecessors Bing provides a more clean and simple interface. As Microsoft covers a major part of operating system market, the general perspective of a user in terms of search engine is that Bing is just another side-product from a technology giant and hence most of them do not take it seriously. But unfortunately it is wrong. Like all the search engines Bing also has some unique features that will force you to use Bing when you need those features. Definitely those features have a unique mark on how we search. We will discuss not only about the special features but also the general operators which can allow us to understand the search engine and its functionalities.

+

This operator works quite similar in all the search engines. This allows a user to forcefully add single or multiple keywords in a search query. Bing will make sure the keywords come after + operator must present in the result pages.

Example: power +search

-

This operator is also known as NOT operator. This is used to exclude something from a set of things, such as excluding a cuisine.

Example: Italian food -pizza

Here Bing will display all the Italian foods available but not pizza. We can write this in another form which can also fetch same result such as the below example

Example: Italian food NOT pizza

“”

This is also same in most of the search engines. This is used to search for exact phrase used inside double quotation.

Example: “How to do Power Searching?”

|

This is also known as OR operator, mostly used for getting result from one of the two keywords or one of the many keywords added with this operator.

Example: ios | android
ios OR android

&

This operator is also known as AND operator. This is the by-default used search operator. If we do nothing and add multiple keywords then Bing will do a AND search in the backend and give us the result.

Example: power AND search
power & search

As this is the default search, it's very important to keep in mind that until and unless we use OR and NOT in capital, Bing won't understand it as operators.

()

This can be called as group operator.

> Grouping of Bing operators supported in following order.
> ()
> “”
> NOT/-
> And/&
> OR/|

As parenthesis has the top priority order, we can add the lower preferred operators such as OR in that and create a group query to execute the lower priority operators first.

Example: android phone AND (nexus OR xperia)

site:

This operator will help to search a particular keyword within a specific website. This operator works quite the same in most of the search engines.

Example: site:owasp.org clickjacking

filetype:

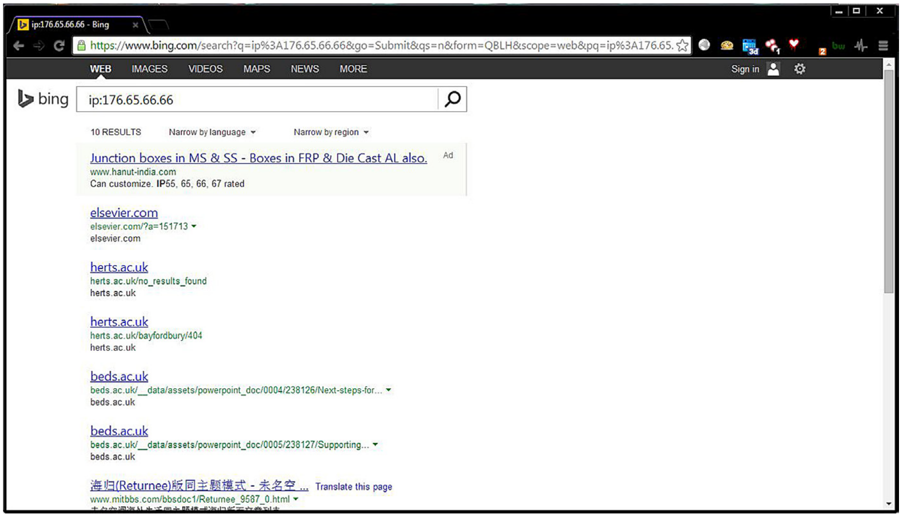
This allows a user to search for data in specific type of file. Bing supports all file types but few, mostly those are supported by Google are also supported by Bing.

Example: hack filetype:pdf

ip:

This unique operator provided by Bing allows us to search web pages based upon IP address. Using it we can perform a reverse IP search, which means it allows us to look for pages hosted on the specified IP).

Example: ip:176.65.66.66



Junction boxes in MS & SS - Boxes in FRP & Die Cast AL also. Ad
www.hanut-india.com
Can customize. **IP**55, 65, 66, 67 rated

elsevier.com
elsevier.com/?a=151713
elsevier.com

herts.ac.uk
herts.ac.uk/no_results_found
herts.ac.uk

herts.ac.uk
herts.ac.uk/bayfordbury/404
herts.ac.uk

beds.ac.uk
beds.ac.uk/__data/assets/powerpoint_doc/0004/238126/Next-steps-for...
beds.ac.uk

beds.ac.uk
beds.ac.uk/__data/assets/powerpoint_doc/0005/238127/Supporting...
beds.ac.uk

海归(Returnee)版同主题模式 - 未名空 ... Translate this page
www.mitbbs.com/bbsdoc1/Returnee_9587_0.html

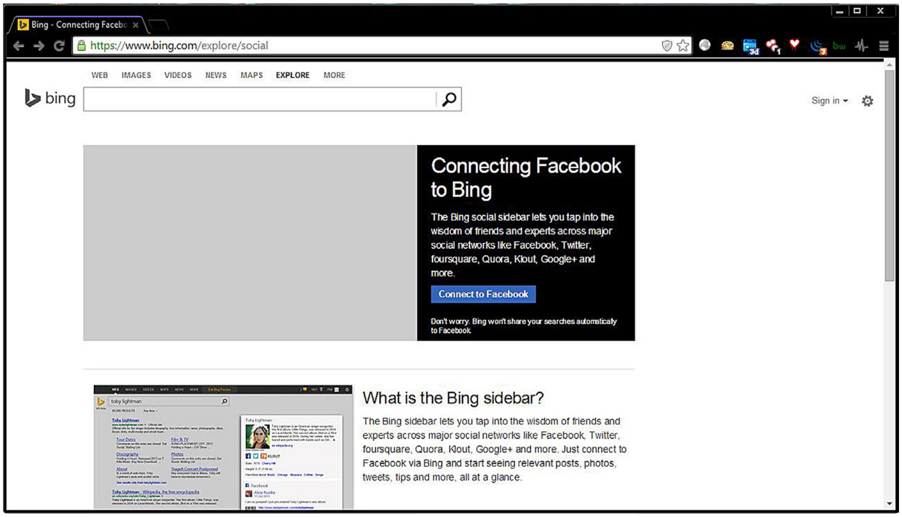
FIGURE 5.4

Bing “ip” search.

feed:

Yet another unique operator provided by Bing is feed, which allows us to look for web feed pages containing the provided keyword.

One other feature that Bing provides is to perform social search using the page https://www.bing.com/explore/social. It allows us to connect our social network accounts with Bing and perform search within them.



Connecting Facebook to Bing

The Bing social sidebar lets you tap into the wisdom of friends and experts across major social networks like Facebook, Twitter, foursquare, Quora, Klout, Google+ and more.

Don't worry. Bing won't share your searches automaticaly to Facebook

What is the Bing sidebar?

The Bing sidebar lets you tap into the wisdom of friends and experts across major social networks like Facebook, Twitter, foursquare, Quora, Klout, Google+ and more. Just connect to Facebook via Bing and start seeing relevant posts, photos, tweets, tips and more, all at a glance.

FIGURE 5.5

Bing social search.

YAHOO

Yahoo is one of the oldest players in the search engine arena and has been quite popular. The search page for Yahoo also has a lot of content such as news, trending topics, weather, financial information, and much more. Earlier Yahoo has utilized third party services to power its search capabilities, later it shifted to become independent and once again has joined forces with Bing for its searching services. Though there is not too much that Yahoo offers in terms of advanced searching as compared to other search engines, the ones provided are worth trying comparing to others. Let's see some of the operators that can be useful.

+

This operator is used to make sure the search results contain the keyword followed by it.

Example: +data

-

Opposite to the "+" operator, this operator is used to exclude any specific keyword from the search results.

Example: -info

OR

This operator allows us to get results for either of the keywords supplied.

Example: data OR info

site:

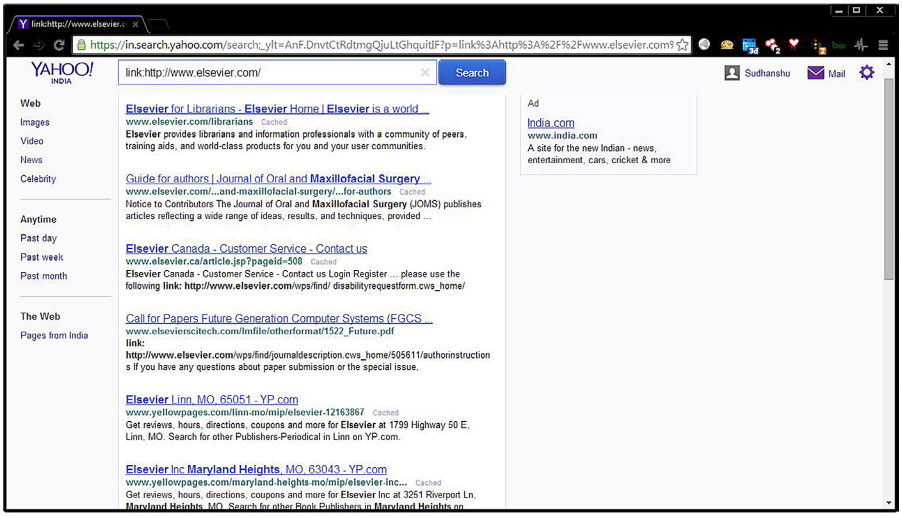
This operator allows restricting the result only to the site provided. We will only get to see the links from the specified website. There are two other operators which work like this operator but do not provide results as accurate or in-depth as they are domain and hostname. Their usage is similar to the "site" operator.

Example: site:elsevier.com

link:

It is another interesting operator which allows us to lookup web pages which link to the specific web page provided. While using this operator do keep in mind to provide the URL with the protocol (http:// or https://).

Example: link:http://www.elsevier.com/


link:http://www.elsevier.c
https://in.search.yahoo.com/search;_ylt=AnF.DnvtCtRdtmgQjuLtGhquitIF?p=link%3Ahttp%3A%2F%2Fwww.elsevier.com
YAHOO! INDIA
link:http://www.elsevier.com/
Search
Sudhanshu
Mail
Web
Images
Video
News
Celebrity
Anytime
Past day
Past week
Past month
The Web
Pages from India

Elsevier for Librarians - Elsevier Home | Elsevier is a world ...
www.elsevier.com/librarians Cached
Elsevier provides librarians and information professionals with a community of peers, training aids, and world-class products for you and your user communities.

Guide for authors | Journal of Oral and Maxillofacial Surgery ...
www.elsevier.com/...and-maxillofacial-surgery/...for-authors Cached
Notice to Contributors The Journal of Oral and Maxillofacial Surgery (JOMS) publishes articles reflecting a wide range of ideas, results, and techniques, provided ...

Elsevier Canada - Customer Service - Contact us
www.elsevier.ca/article.jsp?pageid=508 Cached
Elsevier Canada - Customer Service - Contact us Login Register ... please use the following link: http://www.elsevier.com/wps/find/ disabilityrequestform.cws_home/

Call for Papers Future Generation Computer Systems (FGCS ...
www.elsevierscitech.com/lmfile/otherformat/1522_Future.pdf
link:
http://www.elsevier.com/wps/find/journaldescription.cws_home/505611/authorinstructions If you have any questions about paper submission or the special issue,

Elsevier Linn, MO, 65051 - YP.com
www.yellowpages.com/linn-mo/mip/elsevier-12163867 Cached
Get reviews, hours, directions, coupons and more for Elsevier at 1799 Highway 50 E, Linn, MO. Search for other Publishers-Periodical in Linn on YP.com.

Elsevier Inc Maryland Heights, MO, 63043 - YP.com
www.yellowpages.com/maryland-heights-mo/mip/elsevier-inc... Cached
Get reviews, hours, directions, coupons and more for Elsevier Inc at 3251 Riverport Ln, Maryland Heights, MO. Search for other Book Publishers in Maryland Heights on

Ad
India.com
www.india.com
A site for the new Indian - news, entertainment, cars, cricket & more

FIGURE 5.6

Yahoo “link” search.

define:

We can use this operator to find out the dictionary meaning of a word.

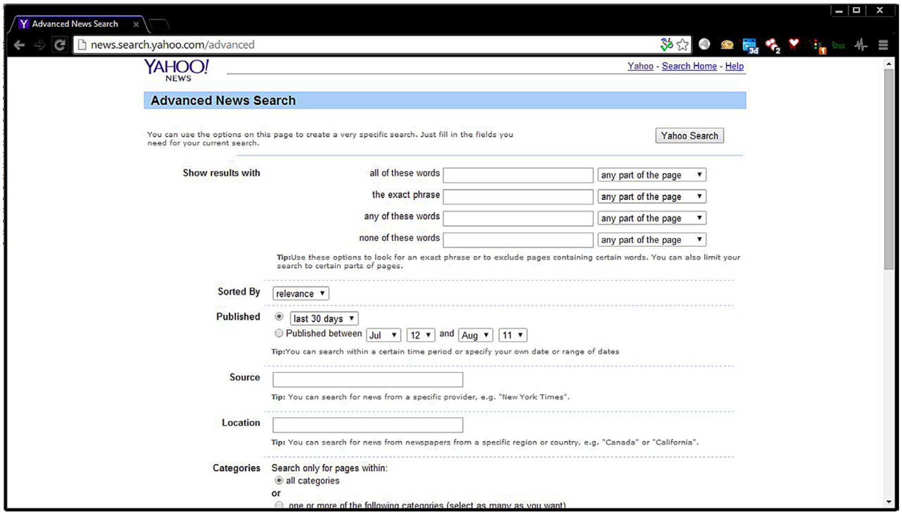
Example: define:data

intitle:

The “intitle” operator is used to get the results which contain the specified keyword in their title tag.

Example: intitle:data

So these are the operators which Yahoo supports. Apart from these we can access the Yahoo advanced search page at http://search.yahoo.com/search/options?fr=fp-top&p=, which allows us to achieve well-filtered search results. One other thing that Yahoo offers is advanced news search which can be performed using the page http://news.search.yahoo.com/advanced .



Advanced News Search

You can use the options on this page to create a very specific search. Just fill in the fields you need for your current search.

Yahoo Search

Show results with

all of these words [ ] any part of the page ▾

the exact phrase [ ] any part of the page ▾

any of these words [ ] any part of the page ▾

none of these words [ ] any part of the page ▾

Tip: Use these options to look for an exact phrase or to exclude pages containing certain words. You can also limit your search to certain parts of pages.

Sorted By: relevance ▾

Published: ● last 30 days ▾

○ Published between Jul ▾ 12 ▾ and Aug ▾ 11 ▾

Tip: You can search within a certain time period or specify your own date or range of dates

Source: [ ]

Tip: You can search for news from a specific provider, e.g. "New York Times".

Location: [ ]

Tip: You can search for news from newspapers from a specific region or country, e.g. "Canada" or "California".

Categories: Search only for pages within:

● all categories

or

○ one or more of the following categories (select as many as you want)

FIGURE 5.7

Yahoo advanced search page.

YANDEX:

Yandex is Russian search engine and is not too much popular outside the country, but it’s one of the most powerful search engines available. Like Google, Bing, Yahoo it has its own unique keywords and data indexed. Yandex is the most popular and widely used search engine in Russia. It’s the fourth largest search engine in the world. Apart from Russia, it is also used in countries like Ukraine, Kazakhstan, Turkey, and Belarus. It is also most under rated search engine as its use is only limited to specific country but in security community we see it otherwise. Most of the people are either happy with their conventional search engine or they think all the internet information is available in the search engine they are using. But the fact is that search engines like Yandex also have many unique features that can provide us with way efficient result as compared to other search engines.

Here we will discuss how Yandex can be a game changer in searching data on internet and how to use it efficiently.

As discussed earlier like other search engines, Yandex has its own operators such as lang, parenthesis, Boolean, and all. Let’s get familiar with these operators and their usage.

+

This operator works quite same for all the search engines. Here also for Yandex, + operator is used to include a keyword in a search result page. The keyword added after + operator is the primary keyword in the search query. The result fetched by the search engine must contain that keyword.

Example: osint +tools

Here the result page might not contain the OSINT keyword but must contain tools keyword. So when we want to focus on a particular keyword or set of keywords in Yandex, we must use + operator.

~~

This is used as NOT operator which is used to exclude a keyword from a search result page. It can be used in excluding a particular thing from a set of the things. Let's say we want to buy mobile phone but not windows phone. Then we can craft a query accordingly to avoid windows phone from search result by using ~~ operator.

Example: mobile phone ~~ windows

~

Unlike ~~ operator ~ is used to exclude a keyword not from search result page but search result sentence. That means we might have both or all the keywords present in the query in a page but the excluded keyword must not be in any sentence with the other keywords mentioned. I understand it being little complicated so let me explain simply. Let's start with the above query

mobile phone ~~ windows

Here if a page contains both mobile phone as well as windows, Yandex will exclude that page from search result.

Example: mobile phone ~ windows

But for the example shown above, it will show all the pages that contains both mobile phone as well as windows but not if these two keywords are in same sentence.

&&

The && operator is used to show pages that contains both the keywords in search result.

Example: power && searching

It will provide the results of all the pages that contain both these keywords.

&

This operator is used to show only pages that contains both the keywords in a sentence. It provides more refined result for both the keywords.

Example: power & searching

/number

It's a special operator which can be used for different purposes according to the number used after slash. It's used for defining the closeness of the keywords. It is quite similar to AROUND operator of Google and NEAR operator of Bing. The number used with slash defines the word distance between two keywords.

Example: power /4 searching

Yandex will make sure that the result page must contain these two keywords with in four words from each other irrespective of keyword position. That means the order in which we created the query with the keywords might change in result page.

What if we need to fix the order? Yes, Yandex has a solution for that also: adding a +sign with the number.

Example: power /+4 searching

By adding the + operator before the number will force Yandex to respond with the results with only pages where these two keywords are in same order and in within 4 word count.

What if we need the reverse of it, let's say we need to get results of keyword "searching" first and after that "power" within 4 word count and not vice versa. In that case negative number will come pretty handy where we can use - sign to reverse what we just did without getting the vice versa result.

Example: power /-4 searching

This will only display pages which contain searching keyword and power after that within 4 word count.

Let's say we want to setup a radius or boundary for a keyword with respect to another; in that case we have to specify that keyword in second position.

Example: power /(-3 +4) searching

Here we are setting up a radius for searching with respect to power. This means that the page is displayed in results shown only if either "searching" will be found within 3 words before or after "power" within 4 word count.

This can be helpful when we are searching for two people's names. In that case we cannot guess that which name will come first and which name will come next so it's better to create a radius for those two names, and the query will serve our purpose.

As we discussed a lot about word-based keyword search, now let's put some light on sentence-based keyword search. For sentence based keyword search we can use Yandex && operator with this number operator.

Example: power && /4 searching

In this case we can get result pages containing these two keywords with in 4 sentence difference irrespective of the position of the keyword. That means either "power" may come first and "searching" after that or vice versa.

!

This operator does something special. And this is one of my favorite keyword. It gives a user freedom to only search a specific keyword without similar word search or extended search and all. What exactly happens in general search is that if you search for a keyword, let's say AND, you will get some results showing only AND and then the results will extend to ANDroid or AMD and so on. If we want to get only result for AND keyword; use this operator.

Example: !and

This will restrict the search engine to provide results only showing pages which contains this particular keyword AND.

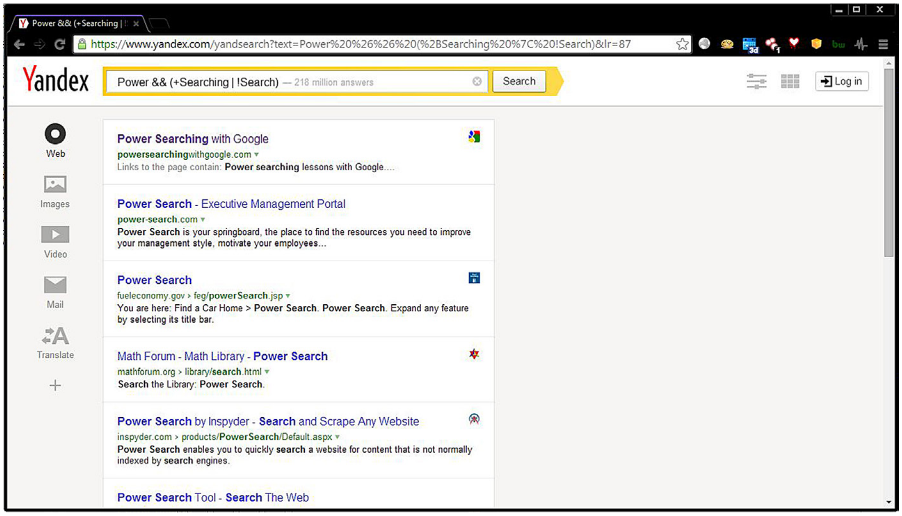
!!

It can be used to search the dictionary form of the keyword.

Example: !!and

()

When we want to create a complex query with different keywords and operators we can use these brackets to group them. As we already used these brackets above, now we will see some other example to understand the true power of this.



https://www.yandex.com/yandsearch?text=Power%20%26%26%20(%2BSearching%20%7C%20!Search)&lr=87

Power && (+Searching | !Search) — 218 million answers

Power Searching with Google
powersearchingwithgoogle.com
Links to the page contain: Power searching lessons with Google....

Power Search - Executive Management Portal
power-search.com
Power Search is your springboard, the place to find the resources you need to improve your management style, motivate your employees...

Power Search
fueleconomy.gov > feg/powerSearch.jsp
You are here: Find a Car Home > Power Search. Power Search. Expand any feature by selecting its title bar.

Math Forum - Math Library - Power Search
mathforum.org > library/search.html
Search the Library: Power Search.

Power Search by Inspyder - Search and Scrape Any Website
inspyder.com > products/PowerSearch/Default.aspx
Power Search enables you to quickly search a website for content that is not normally indexed by search engines.

Power Search Tool - Search The Web

FIGURE 5.8

Yandex complex query.

Example: power && (+searching | !search)

Here the query will search for both sets of keywords first power searching and power search but not both in same result.

""

Now it's about a keyword let's say we want to search a particular string or set of keywords then what to do? Here this operator "" comes for rescue. It is quite similar as Google's "". This will allow a user to search for exact keywords or string which is put inside the double quotes.

Example: "What is OSINT?"

It will search for exact string and if available will give us the result accordingly.

*

This operator can be refereed as wildcard operator. The use of this operator is quite same in most of the search engines. This operator is used to fill the missing keyword or suggest relevant keywords according to the other keywords used in the search query.

Example: osint is * of technology

It will search for auto fill the space where * is used to complete the query with relevant keywords. In this case that can be ocean or treasure or anything. We can also use this operator with double quote to get more efficient and accurate result.

Example: "OSINT is * of technology"

|

This is also quite similar to OR operator of Google. It allows us to go for different keywords where we want results for any of them. In-real time scenario we can search for options using this operator. Let's say I want to buy a laptop and I have different options: in that case this operator will come to picture.

Example: dell | toshiba | macbook

Here we can get result for any of these three options but not all in one result.

<<

This is an unusual operator known as non-ranking "AND." It is basically used to add additional keywords to the list of keywords without impacting the ranking of the website on result. We might not get to know what exactly it does by just going through its definitions. So in simple words it can be used to tag additional keywords to the query list without impacting the page rankings.

Example: power searching << OSINT

It can be used to additionally search for OSINT along with the other two keywords without impacting the page ranking in the result page.

title:

This is quite equivalent to the "intitle." It can be used to search the pages with the keyword (s) specified after title query parameter.

Example: title:osint

This will provide pages that contain OSINT in the title of the web page. Similarly we can use this title query parameter to search for more than one keyword.

Example: title:(power searching)

url:

This "url" search query parameter is also an add-on. It searches for the exact URL provided by the user in Yandex database.

Example: url:http://attacker.in

Here Yandex will provide a result if and only if the URL has been crawled and indexed in its database.

inurl:

It can be used to search for keywords present in a URL or in other words for URL fragment search. This "inurl" query parameter works quite similar in all the search engines.

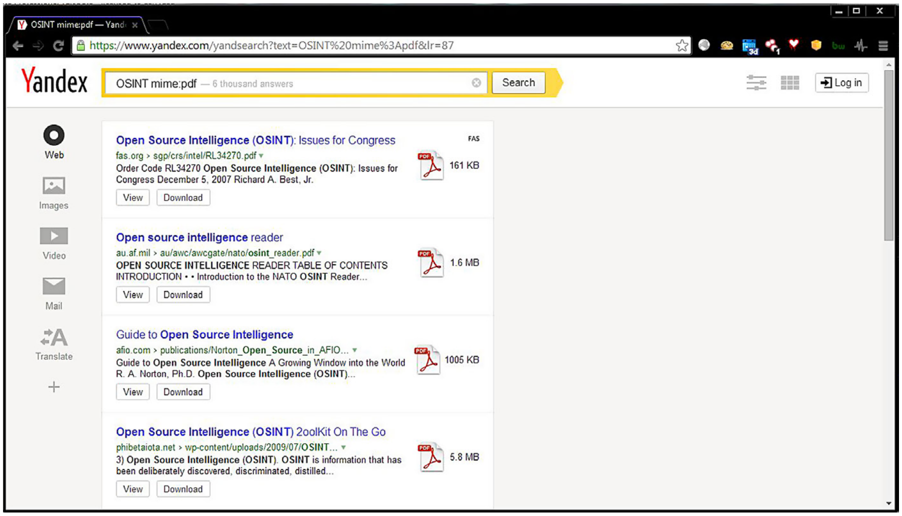
Example: inurl:osint

It will search for all the URLs that contain osint keyword no matter what the position of the keyword is.

mime:filetype

This query parameter is quite similar to "filetype" query parameter of Google. This helps a user to search for a particular file type.

Example: osint mime:pdf


OSINT mime:pdf — Yand
https://www.yandex.com/yandsearch?text=OSINT%20mime%3Apdf&lr=87
Yandex
OSINT mime:pdf — 6 thousand answers
Search
Log in
Web
Images
Video
Mail
Translate

Open Source Intelligence (OSINT): Issues for Congress
fas.org › sgp/crs/intel/RL34270.pdf
Order Code RL34270 Open Source Intelligence (OSINT): Issues for Congress December 5, 2007 Richard A. Best, Jr.
View Download
FAS
161 KB

Open source intelligence reader
au.af.mil › au/awc/awcgate/nato/osint_reader.pdf
OPEN SOURCE INTELLIGENCE READER TABLE OF CONTENTS INTRODUCTION • • Introduction to the NATO OSINT Reader...
View Download
1.6 MB

Guide to Open Source Intelligence
afio.com › publications/Norton_Open_Source_in_AFIO...
Guide to Open Source Intelligence A Growing Window into the World R. A. Norton, Ph.D. Open Source Intelligence (OSINT)...
View Download
1005 KB

Open Source Intelligence (OSINT) 2oolKit On The Go
phibetaiota.net › wp-content/uploads/2009/07/OSINT...
3) Open Source Intelligence (OSINT). OSINT is information that has been deliberately discovered, discriminated, distilled...
View Download
5.8 MB

FIGURE 5.9

Yandex file search.

It will provide us all the PDF links that contains osint keyword. The file types supported by Yandex mime are

PDF, RTF, SWF, DOC, XLS, PPT, DOCX, PPTX, XLSX, ODT, ODS, ODP, ODG

host:

It can be used to search all the available hosts. This can be used by the penetration testers mostly.

Example: host:owasp.org

rhost:

It is quite similar to host but “rhost” searches for reverse hosts. This can also be used by the penetration testers to get all the reverse host details.

It can be used in two ways. One is for subdomains by using the wildcard operator * at the end or another without that.

Example: rhost:org.owasp.*
rhost:org.owasp.www

site:

This operator is like the best friend of a penetration tester or hacker. This is available in most of the search engines. It provides all the details of subdomains of the provided URL.

For penetration testers or hackers finding the right place to search for vulnerability is most important. As in most cases the main sites are much secured as compared to the subdomains, if any operator helps to simplify the process by providing details of the subdomains to any hacker or penetration tester then half work is done. So the importance of this operator is definitely felt in security industry.

Example: site:http://www.owasp.org

It will provide all the available subdomains of the domain owasp.com as well as all the pages.

date:

This query can be used to either limit the search data to a specific date or to specific period by a little enhancement in the query.

Example: date:201408*

In this case, format of date used is YYYYMMDD, but in case of the DD we used wildcard operator “*” so we will get results limited to August 2014.

We can also limit the same to a particular date of the August 2014 by changing a bit in the query.

date:20140808

It will only show results belong to that date.

We can also use "=" in place of ":" and it will still work the same. So the above query can be changed to

date=201408*
date=20140808

As we discussed earlier we can also limit the search results to a particular time period. Let's say we want to search something from a particular date to till date. In that case we can use

date=>20140808

It will provide results from 8th August 2014 to till date, but what if we want to limit both the start date and the end date. In that case also Yandex provide us a provision of providing range.

date=20140808..20140810

Here we will get the results form date 8th August 2014 to 10th August 2014.

domain:

It can be used to specify the search results based of top level domains (TLDs). Mostly this type of the domain search was done to get results from country-specific domains. Let's say we wanted to get the list of CERT-empanelled security service providing company names from different countries. In that case we can search for the country-specific domain extension let's say we want to get these details for New Zealand then its TLD is nz. So we can craft a query like

Example: "cert empanelled company" domain:nz

lang:

It can be used to search pages written in specific languages.

Yandex supports some specific languages such as
RU: Russian
UK: Ukrainian
BE: Belorussian
EN: English
FR: French
DE: German
KK: Kazakh
TT: Tatar
TR: Turkish

Though we can always use Google translator to translate the page from any languages to English or any other languages, it's an added feature provided by Yandex to fulfill minimum requirements of the regions where Yandex is used popularly.

So to search a page we need to provide the short form of the languages.

Example: power searching lang:en

It will search for the pages in English that contains power searching.

cat:

It is also something unique provided by Yandex. Cat stands for category. Yandex categorizes different things based on region id or topic id. Using cat we can search for a result based on region or topic assigned in Yandex database.

The details of Regional codes: http://search.yaca.yandex.ru/geo.c2n.
The details of Topic codes: http://search.yaca.yandex.ru/cat.c2n.

Though the pages contains data in Russian language, we can always use Google translate to serve this purpose.

As we discussed in the beginning that Yandex is an underrated search engine some of its cool features are definitely going to put a mark on our life once we go through this chapter. One of such feature is its advanced search GUI.

There are lazy people like me who want everything in GUI so that they just have to customize everything by providing limited details and selecting some checkbox or radio buttons. Yandex provides that in the below link

http://www.yandex.com/search/advanced?&lr=10558

Here we have to just select what we want and most importantly it covers most of the operators we discussed above. So go to the page, select what you want, and search efficiently using GUI.

Definitely after going through all these operators we can easily feel the impact of the advance search or we can also use the term power search for that. The advance search facilitates a user with faster, efficient, and reliable data in the result. It always reduces our manual efforts to get the desired data. And the content quality is also better in advance search as we limit the search to what we are actually looking for. It can be either country-specific domain search, a particular file type, or content from a specific date. These things cannot be done easily with simple keyword search.

We are in an age where information is everything. Then the reliability factor comes in to picture and if we want bulk of reliable information from the net in very less time span then we need to focus on the advance search. We can use any conventional search engine of our choice. Most of the search engines have quite similar operators to serve the purpose but there are some special features present; so look for those special features and use different search engines for different customized advance search.

So we learned about various search engines and their operators and how to utilize these operators to search better and get precise results. For some operators we say their individual operations and how they can help to narrow down the results and for some we saw how they can be used with other operators to generate a great query which directly gets us to what we want. Though there are some operators for different search engines which work more or less in the same fashion yet as the crawling and indexing techniques of different platforms are different, it is worthwhile to check which one of them provides better results depending upon our requirements. One thing that we need to keep in mind is that the search providers keep on deprecating the operators or features which are not used frequently enough and also some functionalities are not available in some regions.

We saw how easily we can get the results that we actually want with the use of some small but effective techniques. The impact of these techniques is not just limited to finding out the links to websites, but if used creatively they can be implemented in various fields. Apart from finding the information on the web, which certainly is useful for everyone, these techniques can be used to find out details which are profession specific. For example a marketing professional can scale the size of the website of competitor using the operator "site," or a sales professional can find out emails for a company using the wildcard operator "*@randomcompany.com." We also saw how search engine dorks are used by cyber security professionals to find out sensitive and compromising information just by using some simple keywords and operators. The takeaway here is not just to learn about the operators but also about how we can use them creatively in our profession.

We have covered a lot about how to perform searching using different searching platforms in this and some previous chapters. Till now we have mainly focused on browser-based applications or we can say web applications. In the next chapter we will be moving on and learn about various tools which need to be installed as applications and provide us various features for extracting data related to various fields, using various methods.