

IMPROVING THE WEB SEARCH USING SEARCH ENGINES OPERATORS

*Radu G. CREȚULESCU¹, Daniel I. MORARIU¹, Macarie
BREAZU¹*

*¹“Lucian Blaga” University of Sibiu, Engineering Faculty, Computer
Science and Electrical and Electronics Engineering Department*

Abstract

The amount of information stored in the Web is huge and without help from search engines it becomes almost impossible to find what you want. The search engines try to index as many web pages as possible. On April 3rd 2018 there were over 45.5 billion webpages indexed by Google and just over 4 billion web indexed by Bing. Even with the help of search engines, the number of addresses of pages returned by them for a simple query is quite large and we, as people, must look through those pages to find what we are really interested in. Finding the URLs in such a huge data space is challenging and requires some special techniques for refining and reducing the number of returned addresses. In our paper we present some interrogation modifiers and search operators used by the Google search engine to reduce significantly the number of the search results and to increase the quality of that results. The presented techniques and commands for search engines that can be used for finding similar web pages or even to find web pages which repeat information from another sources.

Keywords: Search engines, Google Search operators, similar web pages

1. Introduction

In this paper we will briefly present some information regarding the history of the Internet and the Web. Also we will present some search engines and optimization of search results. We will present some important query modifiers used in search engines like Google and Bing and compare the results obtained using those modifiers.

1 Short history

Searching the Web or the Internet? Which of these is correct? Many users will probably say that “Searching the Internet” is the correct answer. But the Internet (first was called Arpanet) is the name of an older computer network from the 60’ used by the Pentagon together with scientist form MIT and UCLA, the so-called ARPANET (Advanced Research Projects Agency) [WEB_01].

The developing of the Internet protocol begins from 1973 and, by the mid-1970, a nascent Internet was created within the research community. It was not until 1983 that the Internet protocols (TCP/IP) were formally adopted for use on the WEB [WEB_02]. We could say that this was the moment when several networks were interconnected using this protocol and the INTERNET was born. This big network was used for file transferring, remote connections and sending emails (first email was sent in 1971 – by Ray Tomlinson the inventor of the “@”).



Fig. 1 ARPANET – First 4 interconnected computers [WEB_04]

Later, in March 1989 Tim Berners Lee describes the possibility to use linked information systems for access the data in his paper “Document Management System” [Lee_89] which will become the foundation of the WEB. In the paper “WorldWideWeb: Proposal for a HyperText Project” [WEB_05] Tim Berners Lee and Robert Cailliau present the basics of a hypertext document and a structure of a Hypertext Server. You can still find the first webpage at <http://info.cern.ch/hypertext/WWW/TheProject.html>. Berners-Lee’s site went live in 1991, and was accessed by a decidedly small audience of fellow CERN researchers. After launching MOSAIC in 1993 (the first Web-Browser or just Browser, which had included the images tag, so that images could be displayed on the same page with the text) the WEB gets an impressive boost.

So to answer the first question: the correct answer would be that we are searching information on the WEB. We will conclude this section with a quote of Tim Berners Lee: “*The web is more a social creation than a technical one. I designed it for a social effect — to help people work together — and not as a technical toy*”.

2 Searching on the WEB

2.1 The WEB

What is the WEB? The definition of the Web could be: The World Wide Web (WWW) is a distributed hypermedia system which is built on top of some of the services provided by the Internet

In order to find this hyper documents, the user will need an URL (Uniform Resource Locator), a WEB address of the page. Such an address looks like: <http://www.ulbsibiu.ro/bcu/index.php>

The structure of a WEB address is presented in Fig. 2.

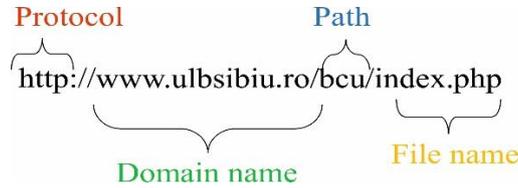


Fig. 2 Structure for the address of a WEB-Pages

To find the information from the WEB you need to know the correct web address, that can be a big problem.

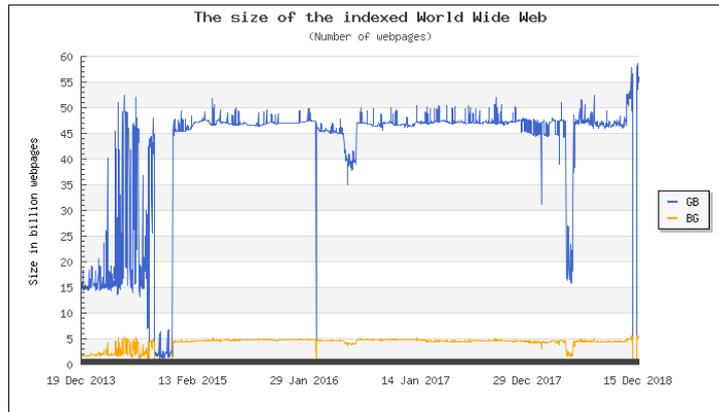
Because of the still growing number of web pages, the user needs some help in discovering information. So that some search engines over the WEB are necessary.

There are several WEB search engines which are used frequently today. In Table 1 we have presented a statistic of most used search engines and the market share [WEB_06]

Table 1 - Search engines:

| No | Search Engine | Share |
|-----|---------------|--------|
| 1. | Google | 73.04% |
| 2. | Baidu | 13.28% |
| 3. | Bing | 7.91% |
| 4. | Yahoo! | 3.70% |
| 5. | Yandex | 1.11% |
| 6. | Ask | 0.42% |
| 7. | DuckDuckGo | 0.24% |
| 8. | Naver | 0.10% |
| 9. | AOL | 0.06% |
| 10. | Dogpile | 0.03% |

In the rest of the article we will compare only the Google and Bing search engines because these are the most used search engines (Baidu is a Chinese search engine and used only for Chinese web pages). For example, the actual number of indexed web pages by the Google and the Bing search engines is given at address www.worldwidewebsite.com. In Fig. 3 we present the number of web pages indexed by the first two search engines (GB sorted on Google and Bing, and BG sorted on Bing and Google).



GB = Sorted on Google and Bing
BG = Sorted on Bing and Google

Fig. 3 Number of WEB-Pages [WEB_03]

2.2 Search engines

Search engines are web applications which automatically search on the web and return several web addresses that contain keywords given by the user. In the back-end of the search engine contains several databases and special programmes called “robots” or “agents” which automatically crawling the Web using the links found on webpages. Each found webpage is indexed (Keywords are extracted) and stored in databases together with the URL.

In the front-end the search engine returns to the user a list with most relevant links (accessing the databases) and a short description called snippets. These snippets contain the title of the page, the URL and a short text which contain the keywords used on performing the search operation. The Google search engine and the Bing search engine are performing the search retrieving URLs based on: title, URL and a part of containing text of webpages.

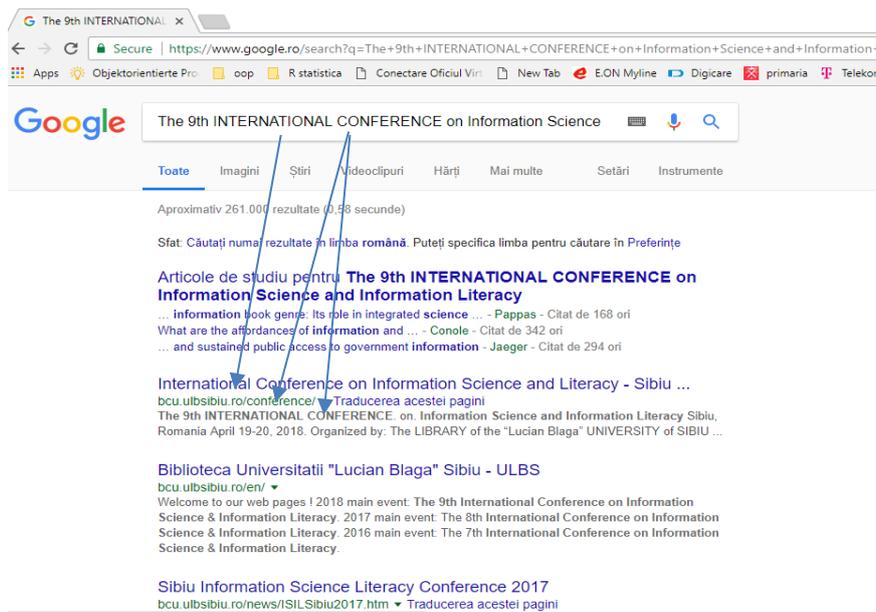


Fig. 4 Search results

The problem to find information on the web is the enormous number of URLs returned by the search engines. If you search after the word BCU (Biblioteca Central Universitară) which is an acronym for the library of the “Lucian Blaga” University of Sibiu, Google will return 7.210.000 results in 0,38 seconds. Now, finding the pages which are relevant to the user needs, is difficult and requires some refining. So that, the actual search engines offer some operators to users in order to reduce the returned pages.

2.3 Search engine operators

In order to find faster URLs that are relevant with the given keyword/keywords the user can use some commands for the search engine called operators. These operators are specific for a given search engine. A complete list for Google operators you can find in [WEB_07] and for Bing operators in [WEB_08].

Next we present some basic operators and some advanced operator that can be used to refine the search.

2.3.1 Basic search operators

In Table 2 we present basic operators and how these operators can be used comparatively between Google and Bing.

Table 2 – Basic operators:

| Google Operator | Bing Operator | Example |
|-----------------|---------------|---|
| " " | " " | "Information Science And Information Literacy" It will force Google to use exact-match. On single words, it excludes synonyms. |
| OR | OR | Morariu OR Cretulescu Google search use as default the AND operator between terms. Specify "OR" for a logical OR operator between words. |
| | | Morariu Cretulescu The pipe () operator is identical to "OR" |
| () | () | (Breazu OR Morariu) Information Science In order to control the order of execution of operators, we can group operators using parentheses. |
| - | - | Cretulescu -Andreea To exclude a term from the results need to put minus (-) in front of any term (including operators) that want to be exclude from the results |
| * | * | "Ultima noapte * razboi" The * sign acts as a wild-card and will match on any word |
| in | | 250 km/h in m/s This operator will make transformations between units |

As we see, both search engines use the same operators for refining basic search operations. By default, both search engines use the AND operator between words and if they find a small number of relevant pages (or not relevant pages) they trays to eliminated some words from user query in order to find some relevant pages.

2.3.2 Advances search operators

In Table 3 we present advanced operators and how these operators can be used comparatively between Google and Bing.

Table 3 – Advanced operators:

| Google Advanced operator | Bing Advanced operators | Example |
|--------------------------|-------------------------|---|
| intitle: | intitle: | intitle: Information Science And Information Literacy Search only in the page's title. Use exact-match (quotes) for phrases. |
| allintitle: | | allintitle: International Conference on Information Science And Information Literacy – Sibiu Search in the page title for every individual term. |
| inurl: | url: | Information Literacy inurl:conference Look for a word or phrase (in quotes) in the web address (URL). |
| allinurl: | | allinurl: Information Science and Information Literacy Search only in the URL for every individual term. |
| intext: | inbody: | intext:"Scientific committee:" Search for a word or phrase (in quotes), but only in the body/document text. |
| allintext: | | allintext: Scientific committee Sibiu April 19-20 2018 Search in the body/ document text for every individual term. |
| filetype: | filetype: | referat Napoleon filetype:doc Match only result that contain a specified file type. Can be used types as PDF, DOC, XLS, PPT, and TXT. |
| site: | site: | "Word *" site:bcu.ulbsibiu.ro Search the term only in the specified web domain |

Some of Advanced operators are look same for both search engines, but Google have more operators that help us to make the query accurate and find faster the relevant URLs. Google search engine permit us to exactly specify where want to be our keywords.

3 Experimental Results

In Table 4 we present the search results performed with basic search operators using the Google and the Bing search engines.

Table 4 – Search results using basic operators:

| Operator | Search string | Results | | Rank of the most useful address | |
|----------|--|----------|----------|---------------------------------|------|
| | | Google | Bing | Google | Bing |
| | Information Science And Information Literacy | 17900000 | 15600000 | 2 | 47 |
| " " | "Information Science And Information Literacy" | 188 | 16 | 1 | 1 |
| - | Cretulescu -Andreea | 111000 | 36700 | 1 | 1 |
| | Ultima noapte razboi | 4150000 | 35500 | 1 | 1 |
| * | "Ultima noapte * razboi" | 2480000 | 15 | 4 | 1 |

The experiments were performed with and without the use of the search operators. As expected, the amount of the returned search results is smaller in case of using the search operators. However, it seems that the google search engine, even not using a search operator returns first the addresses of pages that are relevant for the user. This is caused because the google search engine use also information about the geolocation of the user. So that the web-addresses corresponding to the search query which have the same language as the location of the user and that are in the proximity of the user get a better page ranking and occur at the beginning of the replied list.

In table 5 we present the search results obtained with Google and Bing using advances search operators.

| Advanced operators | String | Results | | Rank of the most useful address | |
|--------------------|---|---------|---------------|---------------------------------|---------------|
| | | Google | Bing | Google | Bing |
| intitle: | intitle: Information Science And Information Literacy | 54600 | 417000 | 4 | 10 |
| allintitle | allintitle:International Conference on Information Science And Information Literacy – Sibiu | 3 | not supported | 1 | not supported |
| allinurl | allinurl: Information Science and Information Literacy | 179000 | not supported | 2 | not supported |
| allintext | allintext: Scientific committee Sibiu April 19-20 2018 | 9700 | not supported | 1 | not supported |
| filetype | referat Napoleon filetype:doc | 1610 | 593 | 1 | 2 |

As it can be observed, not all advanced operators used by Google are supported by Bing Search engine. Therefore, we will focus on the results obtained only by the Google Search engine. As expected, the search results obtained using the advanced operators are better focused on the search term. All searched pages were on rank 1 or 2 that means that the use of this operators could be an advantage. This is a good indicator to use these operators in order to obtain quicker the wanted address.

4 Conclusions

In this paper we have presented some aspects regarding the history of the Internet and of the Web. Also we have compared the results obtained by the Google and Bing search engines using some additional search operators in the search queries and make a comparison between those operators made available by Google and Bing.

In general, we can say that, the use of search operators with Google and Bing search engines returns results that are more relevant whit user needs, regarding the number and quality of the returned Web-addresses.

References

- [Cre_15] Cretulescu, R., Morariu, D, *Dezvoltarea aplicatiilor Web*, Editura Universitatii “Lucian Blaga” Sibiu, 2015, Sibiu
- [Lee_89] Berner-Lee, T, *Information Management: A Proposal*, CERN, 1989, (<https://www.w3.org/History/1989/proposal.html> last accessed June 2018)

- [WEB_01] History of the internet, <https://www.internetsociety.org/internet/history-internet/short-history-of-the-internet/> last accessed June 2018
- [WEB_02] International Telecommunication Union, https://www.itu.int/net/itunews/issues/2010/04/pdf/201004_16.pdf, last accessed June 2018
- [WEB_03] The size of the WWW, <https://www.worldwidewebsize.com/> last accessed June 2018
- [WEB_04] Arpanet, the first internet, <http://classes.design.ucla.edu/Spring06/161A/projects/camile/arpanet/index.html>, accessed in June 2018
- [WEB_05] WorldWideWeb: Proposal for a HyperText Project <https://www.w3.org/Proposal.html>, accessed in June 2018
- [WEB_06] Search Engine Market Share, <https://netmarketshare.com/search-engine-market-share.aspx>, accessed June 2018
- [WEB_07] Google Search Operator, <https://moz.com/learn/seo/search-operators>, last accessed June 2018
- [WEB_08] Bing Advanced search keywords, <http://help.bing.microsoft.com/#apex/18/en-US/10001/-1> last accessed June 2018