International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol. 3, No. 1, 2016, pp. 96-102.
ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

Reviewing the Overlap of Results among Google, Yahoo, and Bing Search Engines

Faranak Salman Mohajer

Master in Computer Engineering, Faculty of Electrical Engineering, Computer and Information Technology, Islamic Azad University, Zanjan, Iran

Abstract

The performance and capabilities of Web search engines has always been one of the major research fields. Every day, millions of people use search engines around the world. This study aimed to review the overlap of results among Google, Yahoo, and Bing search engines. Totally, 100 scientific computer science conferences were searched. From extracted results for any search, the first 10, 20, and 30 links of results were considered. Therefore, 3000, 6000, and 9000 links were extracted from three search engines. However, it was determined that 63% of links were unique in any search engine, 18.4% of links were similar in two search engines, and 0.4% of links were similar in three search engines. According to results, the overlap of search engine results was very low in conference information retrieval. Since the results of tests for first 10, 20, 30 links were very close to each other, it was concluded that the first 20 links are reliable in finding the required information from search engines and give more relevant results to users.

Keywords: Web, Search Engine, Information Retrieval, Overlap, Results Link, Search

1. Introduction

The performance and functionality of web search engines is one of the most popular and notable research issues of scholars and researchers [1]. Today, the search engines are used by indexing billions of web pages by millions of people around the world. Millions of Internet users use search engines to find their needed information. In fact, search engines are an integral part and main reason for using Internet and easiest way to access and read content of web. The Web is a great source of information and provides acceptable answers to majority of users' questions. Today, people consider the internet as the better and fastest way to find the answer of their questions. When users want to find information on the Internet, the use of search engines will be their first step. Therefore, the abilities, capabilities, limitations, and problems of search engines are very important.

When the user enters the keyword into a search engine, the results are immediately displayed to user on a regular basis. These results are arranged so that the most relevant results to user requests have higher rating. It is noteworthy that search engines do not always provide accurate results; however, each search engine uses its own algorithms to find the results as much as possible and rank them. These algorithms are specific and secret in any search engine. For this reason, different results are obtained from entering the same word.

The search engines overlap means provision of similar results by search engines. For example, if same keyword is entered to three search engines and first 20 results in any search engine are retrieved, the shared results between two search engines and three search engines may be extracted and the overlap of them may be investigated. One of the common problems of users is poor overlap of search engines to find users' query. In other words, when user enters the keyword in a search engine and different results are provided in different search engines, the user will get confused about which of the search engine results are more accurate and more appropriate response to his/her query.

Ding and Marchionini (1996) investigated the first results of three search engines including Infoseek, Lycos, and OpenText with five inquiry. They found that there is very little overlap among these search engines. Lawrence and Giles (1998) also showed that any search engine do not index more than 16% of all web sites. These studies began research on differences among search engines in indexing, retrieval, and rankings of results. In fact, search engines have three major difference: frequency of updates, indexing, and ranking of information [3, 4].

Baharat and Broder (1998) used the random searches to assess overlap among search engines. They showed that the overlap among Hotbot, AltaVista, Excite, and Infoseek is 1.4% [5].

Esfandiari Moghadam and Bahari Movafagh (2012) investigated four search engines: Yahoo, Live Search, Google, and Ask. In this study, 10 queries in medical field were used. The findings showed that Yahoo has the highest unique results and Google has the most overlap with other search engines. Also, the overlap among four search engines was 11% [14].

This study investigates the overlap of search engines by entering same queries to several search engines and retrieving information from them to determine differences and similarities in search engines' retrieved data.

2. Methodology:

In this study, the three major web search engines were selected: Google, Yahoo, and Bing. According to Table 1, they are the most used search engines at present (97.3%) (ComScore qSearch Data, November 2015) [20].

Table 1: The percentage of using search engines

Search engine	October 2015	November 2015
Google	% 63.9	% 63.9
Bing	% 20.8	% 20.9
Yahoo	% 12.5	% 12.5
Others	% 2.8	% 2.7

In this study, the scientific conferences in the field of computer science were searched. Since the name of most scientific conferences is an acronym, therefore, the acronym of conferences' names was searched. The DBLP (Digital Bibliography and Library Project) database was used to extract the conference names. The DBLP is a database of computer authors and scientists and has a set of research records in computer science. It has been collected at University of Tier and has around 3 million records (DBLP, July 2015) [17]. This database is a popular tool for following computer science colleagues and researchers. It is also a means to retrieve information and detailed records for use in new scientific research [15].

The DBLP data can be downloaded on the Internet in XML file format. This is a very large collection of research records in the world. In many scientific articles, numerous researchers have used this database to test their algorithms.

The number of queries was 100 conference names which were extracted from DBLP database.

The SEKE (**International Conference on Software Engineering and Knowledge Engineering**), ICSE (**International Conference on Software Engineering**), and ICCSIT (**International Conference on Computer Science and Information Technology**) are some of the queries in this study.

It should be noted that after extraction of 100 queries from DBLP database and putting them in a new database, the designed application in this study entered each of inquiries into three search engines. Then, the results links of search engines were recorded in database. According to groupings, the links were compared with each other and the number of unique links and common links were determined separately.

3. Analysis of data

Providing 100 queries to each of search engines and considering Group A 10 first results links (as shown in Table 2), the number of extracted links for total queries was 1,000 links, for Group B was 2000 links, and for Group C was 3000 links. If three search engines are considered for each group, there will be 3000, 6000, and 9000 links for groups A, B, and C, respectively. A total of 18,000 links were extracted and each query was compared in each group.

Table 2: Number of inquiries and extracted links from search engines

Name of groups	Number of inquiries	Extracted links for all inquiries	Total extracted links from three search engines
Group A (10 first link)	100	1000	3000
Group B (20 first link)	100	2000	6000
Group C (30 first link)	100	3000	9000

After extracting all links, they should be compared to determine the overlap of links in search engines.

4.1: Group A (10 first results)

Extracting first 10 results links for each query and comparing the links in search engines, it was determined that from 1000 extracted links, Google search engine had 845 unique links, Yahoo had 527 unique links, and Bing had 500 unique links. According to Table 4, the comparison of links in two Google and Bing search engines shows that only 91 links from 1000 links were shared in these two search engines. For Google and Yahoo search engines, this number was 64 links and for Yahoo and Bing, it was

409 links. Finally, comparing the links of three search engines, it was shown that from 1000 links, only 45 links were common in three search engines.

Table 3. Number and percentage of extracted unique links in Group A

Search engine	Number of unique links	Percentage of uniqueness
Google	845	%85
Yahoo	527	%53
Bing	500	%50

Table 4. Number and percentage of shared extracted links in Group A

Search engine	Number of shared links	Percentage of shared links
Google & Bing	91	%9
Yahoo & Bing	409	%41
Google & Yahoo	64	%6
Google & Yahoo & Bing	45	%5

4-2: Group B (20 first results)

Extracting the first 20 links for each query and comparing them (tables 5 and 6), it was shown that from 2000 extracted links in Google search engine, 1743 links were unique. The shared links between Google and Yahoo search engines was only 124 links; the number of shared links among three search engines was 76 links.

Table 5. Number and percentage of extracted unique links in Group B

Search engine	Number of unique links	Percentage of uniqueness
Google	1743	%87
Yahoo	1028	%51
Bing	1019	%51

Table 6: Number and percentage of extracted common links in Group B

Search engine	Number of shared links	Percentage of shared links
Google & Bing	133	%7
Yahoo & Bing	848	%42
Google & Yahoo	124	%6
Google & Yahoo & Bing	76	%4

4-3: Group C (30 first results)

In this case, from 3000 extracted links in any search engine (Table 7 and 8), 2603 links were unique in Google search engine. In Google and Yahoo search engines, the number of shared links was 197 links; in all three search engines, only 100 links were common.

Table 7. Number and percentage of extracted unique links in Group C

Search engine	Number of unique links	Percentage of uniqueness
Google	2603	%87

Yahoo	1493	%50
Bing	1490	%50

Table 8. Number and percentage of extracted common links in Group C

Search engine	Number of shared links	Percentage of shared links
Google & Bing	200	%7
Yahoo & Bing	1310	%44
Google & Yahoo	197	%7
Google & Yahoo & Bing	100	%3

4.4: Comparison of results in three groups

According to Figures 1 and 2, the percentage of unique links in each of the Google, Yahoo, and Bing search engines is very high. Also, the percentage of shared links in Google and Yahoo, Google and Bing, Yahoo and Bing, and among three search engines is very low compared to unique links.

Calculating the unique links in three search engines and shared links between two search engines and among three search engines in each group (Tables 9, 10, and 11), it can be seen that the number of unique links in each of three search engines is 1872 links from 3000 links in group A, 3790 links from 6000 links in group B, and 5586 links from 9000 links in group C. Also in two search engines, 564 links from 3000 links was for Group A, 1105 links from 6000 links was for Group B, and 1707 links from 9000 links was for Group C. And finally in three search engines, only 45 links from 3,000 links was for Group A, 76 links from 6000 links was for Group B, and 100 links from 9000 links was for Group C.

Table 9. Number and percentage of unique and shared results in search engines in Group A

Search engine	Number of unique links	Number of shared links	Percentage of unique links	Percentage of shared links
One search engine	1872		62.4	
Two search engines		564		18.8
Three search engines		45		5

Table 10. Number and percentage of unique and shared results in search engines in Group B

Search engine	Number of unique links	Number of shared links	Percentage of unique links	Percentage of shared links
One search engine	3790		63.1	
Two search engines		1105		18.4
Three search engines		76		4

Table 11. Number and percentage of unique and shared results in search engines in Group C

Search engine	Number of unique links	Number of shared links	Percentage of unique links	Percentage of shared links
One search engine	5586		62.1	
Two search engines		1707		18.9
Three search engines		100		3

According to Figure 3, the percentage of unique links in search engines is high. The percentage of shared links in two search engines is low. The percentage of shared links in three search engines is minimal.

4. Conclusion:

According to results, the number of unique links in three Google, Yahoo, and Bing search engines for queries provided to each of the search engines is very high. The number of shared links between Google and Yahoo as well as Google and Bing is very low. Also, the number of shared links among three search engines is very low. The shared links between Yahoo and Bing is more than between Google and other search engines. This indicates the similarity of search algorithm in these two search engines. However, this is low compared with the number of unique links in each of these search engines. Therefore, it can be concluded that the overlap among search engines results links is very low.

Comparing the results obtained in three groups (Figure 1 and 2), it is quite evident that the results are very similar to each other. This means that the percentage of unique links in 10 first links is not different from 20 and 30 first links. Also, the overlap among three search engines is the same in 10, 20, and 30 first results links. Therefore, it can be said that by increasing the number of links provided by search engines, there will no difference in results; because the most relevant results are often provided at first links. Therefore, the first 20 links may be used to evaluate the overlap.

References:

1- Spink Amanda, J.Jansen Bernard, Blakely Chris, Koshman Sherry, " A study of results overlap and uniqueness among major Web search engines ", Information Processing and Management , v42 (2006), pp 1379–1391

2- Cambazoglu B. Barla, "Review of "Search Engines: Information Retrieval in Practice" by Croft, Metzler and Strohman", Information Processing and Management, v46 (2010), pp 377–379

3- Ding Wei, Marchionini Gary, "A Comparative Study of Web Search Service Performance", Proceedings of the ASIS Annual Meeting, v33 (1996), pp136-142

4- Lawrence Steve, Giles C. Lee, "Searching the world wide web", Science, v 280, no 5360 (1998), pp 98-100

5- Bharat Krishna, Broder Andrei, "Technique for measuring the relative size and overlap of public Web search engines", Computer Networks and ISDN Systems, v30 (1998), pp 379-388

6- Bar-Ilan Judit, Mat-Hassan Mazlita, Levene Mark, "Methods for comparing rankings of search engine results", Computer Networks, v50 (2006), pp 1448–1463

7- Lewandowski Dirk, "The retrieval effectiveness of web search engines: considering results descriptions", Journal of Documentation, v 64 No 6(2008), pp 915-937

8- Rather Rafiq Ahmad, Lone Fayaz Ahmad, and Shah Gulam Jeelani, "Overlap in Web Search Results: A Study of Five Search Engines", Library Philosophy and Practice (e-journal), (2008), Paper 226

9- Usmani Tauqeer Ahmad, Pant Durgesh, Bhatt Ashutosh Kumar, "A Comparative Study of Google and Bing Search Engines in Context of Precision and Relative Recall Parameter", International Journal on Computer Science and Engineering (IJCSE), v4 (2012), pp 21-34

10-Lewandowski Dirk , "Evaluating the retrieval effectiveness of Web search engines using a representative query sample ", Journal of the Association for Information Science and Technology, v 66, Issue 9 (2015), pp 1763–1775

11-Kumar Kailash, Bhadu Vinod, " A COMPARATIVE STUDY OF BYG SEARCH ENGINES", American Journal of Engineering Research (AJER), v2, Issue 4 (2013), pp-39-43

12-AJAYI Olusola Olajide, ELEGBELEYE Damilola Matthew, " Performance Evaluation of Selected Search Engines ", Computer Engineering and Intelligent Systems, v5, no1(2014)

13-Sterling Greg, "Different Engines, Different Results", A Research Study by Dogpile.com, http://searchengineland.com/dogpile-search-results-on-major-engines-diverging-11364, (1394/9/17)

14-Isfandyari Moghadam Alireza, Bahari Movaffagh Zohreh, " The Overlap Rate Of Searching Medical Keyword In General Search Engines", Health Information Management, v9, no 2 (2012), pp 2003 – 2014

15-Alwahaishi Saleh, Martinovic Jan, Snasel Vaclav, Kudelka Milos, "Analysis of the DBLP Publication Classication Using Concept Lattices", Dateso (2011), pp. 132-139

16-J. Jansen Bernard, Spink Amanda," How are we searching the World Wide Web? A comparison of nine search engine transaction logs", Information Processing and Management, v42 (2006), pp 248–263