

# An Analysis of Quality of Freely Available Search Engines

M. Ramanan \*

\*Department of Interdisciplinary, University College of Jaffna, Sri Lanka

Email: [mmramanan@yahoo.com](mailto:mmramanan@yahoo.com)

## Abstract:

Due to the enormous size of the World Wide Web (WWW), search engines are the common means to find web pages. However, there are few numbers of freely available search engines among hundreds and the count keeps increasing as a result of the inspiration of internet advancement. The research would help the users to identify the overall status of the search engines. For this study, seven well-known search engines were elected and compared. The analysis was performed through an experimental method. To make the experiment more effective, sample queries were executed in the selected search engines and the results obtained from them were recorded. With the responses of the queries from the search engines, further analysis has been conducted to evaluate the quality comparison among them. To achieve this, the search-specific formula known as Precision and Mean Reciprocal Rank were used.

### Keywords:

Search engines, Internet, Precision, Mean Reciprocal Rank

## I. INTRODUCTION

The Internet has become the largest available source to obtain the required information in any category. In addition, the main reason why people go to Internet for browsing is for finding the relevant information that they are looking for. However, the only problem is to determine the relevant websites that are containing the essential information among the rest of trillions (or even more as the size is still unknown because it is growing everyday) of websites[1]. Fortunately, there are several search engines available to facilitate the user to get the desired information. As a result, searching and getting the required information from the internet has become simpler [2].

### A. Search Engine

The search engine is a specific software application whose functionality is searching the websites based on the user given words as search [3]. Therefore, it accepts the search requests as input, processes them, and delivers the results based on the request. However, it is a very complicated procedure. To achieve this, every search engine uses a highly complex software system which contains complicated [4].

### B. How does Search Engine Work

Search engines consist of extremely in-depth procedures and methodologies [5]. Also, it is required to be updated all the time due to advance technological

development and upcoming technologies. Because different search engines use a different methodology, it is obvious that each search engine results are different. Search engines scan millions of websites and store the information in their database. When the user produces a search query, the search engine quickly processing the database to get the relevant results of the user request. Search engine ranks the results and sorts based on relevancy and finally, the user gets the results. Crawler, Indexer, and Query handler are key mechanisms used by search engines to retrieve information.

### C. Quality of Search Engines

A search engine quality, ideally, will be considered to have following major aspects.

- Search results are up-to-date
- Search results are relevant
- Most relevant result is in top of the first page.
- Search speed is high

### D. Objectives

The major aim of this research is to analyze the quality attributes of the freely available search engines and comparing the efficiency of them. Accordingly, to fulfill the research objective it is mandatory to consider all of them. Therefore, to make this research to be more effective, a list of seven vitally used search engines is derived after performing some initial investigations. The selected search engines are given below.

- Google
- Bing
- Yahoo
- AltaVista
- AOL search
- Lycos
- Ask jeeves

## II .LITERATURE REVIEW

As search engines started to appear in the internet world, the research comparing search engines had started by the researchers. The early research done by Ding and Marchionini [6] included only three search engines for their research and they discovered that there were no statistically significant differences between them in terms of effectiveness. The same study made by Nicholson on the basis of chaos theory and also produced the same short of result [7].

Another testing was made by Tomaiuolo and Packer with 200 queries and the first 10 responses are been taken too collected [8]. This is also made on the three search engines know as Lycos, InfoSeek, and AltaVista. It also includes another two more human evaluative search tools.

Next year, Clarke and Willett performed research on searches by considering the three different search engines: AltaVista, Lycos, and Excite [9]. Their finding was that there were actually considerable differences between AltaVista and Lycos, and between AltaVista and Excite, however, there were not much significant differences between Lycos and Excite. They concluded that the AltaVista system performed in high level. The reason behind this is that Lycos and Excite search engines were using the technology of extract search algorithm while AltaVista used a different approach known as a full-text search algorithm.

Leighton and Srivastava compare the search with 15 queries on 5 search engines [10]. The researchers resolute the importance of themselves and by completing the cloaking that they identify the search engine of that retrieved the links. Gordond and Pathak selected twelve search engines that were available at that time [11]. Their research target was to identify the three engines (out of twelve) that are satisfying “appropriate experimental design and evaluation”. The presented comparison outcomes by them were considered as one of the most comprehensive at that period. However, Gordon and Pathak stated that the search effectiveness was generally low and there were considerable differences between engines. To some extent, Gordon and Pathak had to be strict of the relevance criterion to rank the engines.

Su performed research on four (Lycos, AltaVista, Infoseek, and Excite) search engines [12]. The researcher requested thirty-six students to obtain and run the students’

individual interest of search queries in all the search engines. The researcher used the first twenty links of the results. To evaluate, relevancy, user satisfaction, efficiency, utility, and connectivity were used as the performance measurements. The result of the research was AltaVista delivered the maximum precision among all. Next year, Wu and Li performed research on four search engines (Google, AltaVista, HotBot, and AllTheWeb) [13]. According to the experimental results, the performances of the selected search engines are nearly equal.

## III. RESEARCH METHODOLOGIES

The research methodology illustrates how the research is proposed to enhance to achieve the research objective. As a result, the research plans are set to collect information from various sources (books, journals, internet, etc...) that are available to conclude how the previous researches conducted when they evaluated the search engines. In addition, as the research principle is also focused on analyzing the quality of the search engine search results, the objective path is set to deal with them as well. After analyzing the previous studies, the experimental research method is considered for evaluating the quality of the search engines. To make the process successful, a set of queries, which are strictly user-oriented as it is one of the research objectives, should be created and the results should be analyzed in graphs.

### A. Data Collection

The study will consist of executing some sample search queries in the search engine and gather the outcomes. It has been initiated with fifteen sentences in “natural language” queries. The queries are constructed either based on the likelihood of covering a wide range of aspects or based on the words used to contain meaning. Subsequently, the queries would be run in all the considered search engines.

## IV. EXPERIMENT

### A. Precision

The term precisions defined as the count of the relevant documents given by the search engine out of the total count of documents given by that search engine. Therefore if the value of precision is 1.0 then it means every result is shown for the search, certainly belongs to the relevant category.

In practice, search hits given by the search engine are often very high. Therefore, it would be extremely difficult to view all the hits to get the precision value. As a result, cut-off rank has been introduced, Based on this approach, precision at one document retrieved (often called as P@1) will be defined as one divided by the total count of documents found until detecting the first relevant document. For instance, if the fifth document is the first relevant document, then P@1 is equal to 1/5 which will be 0.2. Similarly, precision at two documents retrieved (named

as P@2) will be defined as two divided by the total number of documents found until the second relevant document. For example, if the eighth document is the second relevant document, then P@2 is equal to 2/8 which will be 0.25.

Normally the first relevant document would be retrieved very quickly. However, the subsequent relevant documents may appear a bit later. Therefore to decide the quality of the search engine P@1 is not the ideal measurement. In this case, it would be reasonably good when considering P@5. However, there is much better measurement than this. That is taking the average precision values of all the cut-offs from 1 to 5. Ultimately, this will be stated in the below equation.

Where P@ n = precision at n documents recovered

P@1.5 = the average of precision values at every level cutoff from 1 to 5.

This is useful when evaluating the accuracy more detailed in a single query. Besides, P@1.5 will be used for comparing the search engine results in graph format.

**B. Mean Reciprocal Rank**

The mean reciprocal rank is another measurement and it is defined as the average of precision at specific number of documents for the set of queries. For example, if precision at one document is considered for three different queries, then, mean reciprocal rank with cut-off one is given below.

$$MRR_1 = \frac{(P@1 \text{ of query1} + P@1 \text{ of query2} + P@1 \text{ of query3})}{3}$$

$$MRR_2 = \frac{(P@2 \text{ of query1} + P@2 \text{ of query2} + P@2 \text{ of query3})}{3}$$

Where,

P@n = precision at n documents recovered

MRR<sub>1</sub> = Mean Reciprocal Rank of first relevant document

MRR<sub>2</sub>=Mean Reciprocal Rank of second relevant document

Table 1. SUMMARY OF MEAN RECIPROCAL RANK

Search Engine	Google	Bing	AOL	Yahoo	Lycos	Ask	Alta Vista
Mean Reciprocal Rank (MRR) /Average per query (P@1)	0.867	0.956	0.900	0.717	0.811	0.836	0.733
Mean Reciprocal Rank (MRR) /Average per query (P@1.5)	0.758	0.863	0.747	0.715	0.753	0.733	0.737

**V. ANALYSIS**

As indicated in the introduction the search engines are compared with each other. To achieve this, the features of the ideal search system would be used. The ideal search engine would have the following features that are listed below.

- Search results are up-to-date
- Search results are relevant
- Most relevant result is in top of the first page.

**A. Search Results are Up-to-Date**

It is essential for search engines to deliver the up-to-date results for the search queries of the users. This ensures that the specific search engine is correctly identify the sites and make the results are updated frequently.

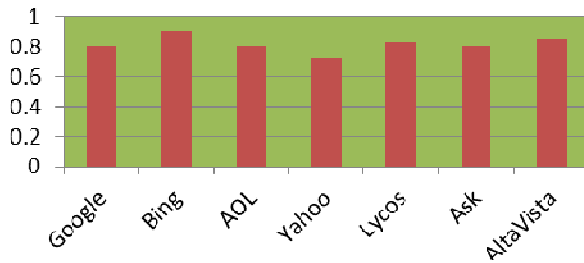


Fig.1: Compare P@1.5 for "query no 1"

When comparing the results, all except Yahoo are able to reach or pass 0.800 which would be truly appreciated by the users. Bing has shown its potential on this by reaching 0.900. AltaVista receives second place while Lycos is in third place. The market leader Google and AOL, both are lying in fourth place.

**B. Search Results are Relevant**

It is vital to give the results that are relevant for the search queries of the users. To measure whether the results are relevant, Mean Reciprocal Rank (MRR) would be used and it would indicate the position of the search engines. The following graph is comparing the MRR values of selected search engines.

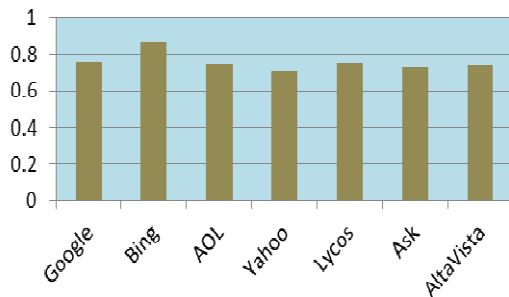


Fig 2. Comparing MRR1.5 of search engines

The above graph is obtained by considering the mean reciprocal rank of the average of all the five precision (MRR1.5). As per the graph, Bing achieved first place. Although Yahoo shows as it is in the lowest level, however, the remaining search engines are nearly lying in the line 0.750. Using the values, it can be said that Google is in second place while Lycos is in third place.

**C. Most Relevant Result in Top of the First Page.**

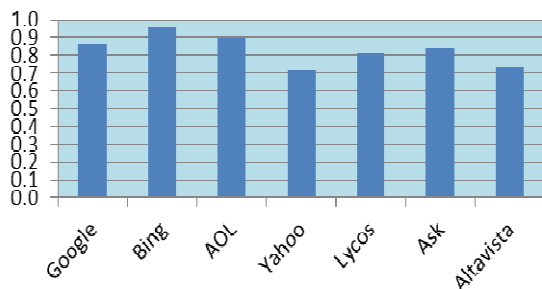


Fig 3. Comparing MRR1

When comparing MRR1 values of the search engines, Bing is holding the highest value and Yahoo holds the lowest value. AOL holds the second-highest and Google is in third place. Although Google has been used by the majority of the people and got a wide reputation, the latest search Bing has gained first place while considering the MRR1. Furthermore, except Yahoo and AltaVista, the rest of the five search engines are reached 0.800. However, all the selected engines are given more than 0.700.

**VI. CONCLUSION**

For the experiment seven search engines were elected, and fifteen queries were developed and independently executed on each of the selected search engines. The query results returned for these queries were analyzed. However, none of the search engines has given relevant results as its top most for all the queries. In this research, the well-known search formula called Precision and Mean Reciprocal Rank were used to analyze the search results.

Based on the results, when considering Google, it performed excellently in queries related to such as hotels, weather, software, education and health; however, it lacks in the area like Flight information, and the latest products. For Bing, the strongest areas are the latest products, software, weather, movies, health, and hotels; thus, its weakest areas are online books and a special offer on items. As AOL web search is powered by a Google search engine, its strengths and weaknesses are similar to Google. Lycos search engine shows its potential in subjects like travel, hotels, and education; however, it extremely struggles in areas like books and special offers on items. Yahoo and AltaVista both show their capabilities in hotels, education, food, and Travel; thus, they both lack in books, special offers, and jobs. Finally, the Ask search engine shows valuable results in areas similar to hotels, foods, jobs, and software, while it fails in categories like special offers, latest products, and books

By this research, Bing might be ranked as top among all the search engines. However, it should be noted that when looking at the history, different search engines obtained higher ranks for different research.

**REFERENCES**

- [1] Levene M, "An Introduction to Search Engines and Web Navigation", second edition, John Wiley and Sons, Inc., pages: 13-14, 2010
- [2] Franklin C, (nd), how internet Search engine Work, web page available at <http://computer.howstuffworks.com/internet/basics/search-engine.htm> last accessed on 19/12/2020
- [3]Schlein A.M, Find it online: The complete Guide to online research, Fact on demand press, 4th edition, 2004
- [4] Meghabghab G, Kandel A, "Search Engines, Link Analysis, and User's Web Behavior", Springer pages 84-85 , 2008
- [5] Boswell W, how does a search engine work? The Basic Inner Workings of Search Engines, web page available

at:<http://websearch.about.com/od/enginesanddirectories/a/searchengine.htm> last accessed on 29/11/2020

- [6] Marchionini G, “Comparative study of web search service performance”, In Proceedings of the ASIS Annual Conference ,1996
- [7] Nicholson S. “Raising reliability of Web search tool research through replication and chaos theory”. Journal of the American Society for Information Science, 51(8), 724–729 ,2000
- [8] Tomaiuolo N, Packer J. “An analysis of Internet search engines: assessment of over 200 search queries” Computers in Libraries, 16(6), 58–62, 1996
- [9] Clarke S.J, Willett P, “Search Engine Precision and Recall” web page available at [http://www.dba-oracle.com/t\\_search\\_engine\\_precision\\_recall.htm](http://www.dba-oracle.com/t_search_engine_precision_recall.htm) last accessed on 18/12/2020
- [10] Leighton H. V, Srivastava J, “First 20 precision among World Wide Web search services (search engines)”. Journal of the American Society for Information Science, 870–881, 1999
- [11] Gordon M, Pathak P, finding information on the World Wide Web, The retrieval effectiveness of search engines, Information Processing and Management pages 141-180, 1999
- [12] Su L. T , “A comprehensive and systematic model of user evaluation of Web search engines”. Journal of the American Society of Information Science and Technology, pp 12-18, 2003
- [13] Wu S. L, Li J.” Effectiveness evaluation and comparison of Web search engines and meta-search engines”. Advances in Web-Age Information Management: 5th International Conference,303–314, 2004
- [14] Brin S, Page L, Google’s search architecture: Peak Positions Search engine optimization web page available at <http://www.peakpositions.com/how-google-works-google-keyword-search-system.html> last accessed on 08/12/2020