

EVALUATING SEARCH EFFECTIVENESS OF SOME SELECTED SEARCH ENGINES

M. S. BUTE, ADAMU HUSSAINI AND INUWA HADIZA ADAMU

(Received 19 September 2016; Revision Accepted 4 October 2016)

ABSTRACT

With advancement in technology, many individuals are getting familiar with the internet a lot of users seek for information on the World Wide Web (WWW) using variety of search engines. This research work evaluates the retrieval effectiveness of Google, Yahoo, Bing, AOL and Baidu. Precision, relative recall and response time were considered for this evaluation, a total of 24 search queries were sampled based on information queries, navigational queries and transactional queries. They are categorized into single word queries, double word queries, sentence queries and alphanumeric queries. Finally the overall average of all the designated queries shows Bing has the best score based on relative recall with 0.54, Google has the best response time with 0.39s and has the highest precision with 1.37.

KEYWORDS: World Wide Web, Search Engine, Search Query, Links.

INTRODUCTION

The primary objective of information retrieving system (IRS) is to find or access relevant documents with respect to user's query. The documents are being retrieved as a ranked list, whereby the ranking is estimated based on relevance. Today, while searching for information on World Wide Web (WWW), one usually carries out a query request through a term based search engine. The World Wide Web is a rapidly expanding hyperlinked collection of unstructured information, a friendly user interface and also contains a hypermedia features for which have been attracting number of information providers Kumar and Bhadu (2013).

The growth of the World Wide Web (WWW) is an unprecedented trend. Four years after the web birth in 1990, millions or more copies of the well known Web browser Mosaic was in use. The growth was as a result of increase of the web servers, value and the number of the web pages that are accessible by these servers. In 1999 the number of Web servers was estimated at about 3 million and the number of Web pages at about 800 million, and three years later, in June 2002, the search engine AlltheWeb announced that its index contained information about 2.1 billion Web pages Can, Nuray and Sevdik (2004).

A web search engine is an information retrieval system, which is used to locate Web pages relevant to user queries. The search engines do not really search Web directly, each one searches a database of the full text of Web pages selected from the billions of web pages. Search engine databases are built by a computer robot programs called Spider. They are responsible for finding

the pages for potential inclusion by following the links in the pages they already have in their database. Also, search engine spiders cannot find a Web page that is never linked to any other page. The only way such can get into the search engine companies is by requesting that the new page be included. After the spiders find the pages, they pass them to another computer program for indexing. It is used for indentifying text, links and other content in the page and store it in the search engine database's files so that the database can be searched by keyword Akinola (2010).

The effectiveness of search engines depend on the features it has. Some search engines retrieve document within few seconds. The Web crawler is responsible for identifying the data available in the vast sphere. Web crawling is carried out over web search engines recursively to offer up to date data to the users. The arrangement of Web page also plays a vital role, whereby it matches the user query with the Web page existing in the database. Website indexing searches for specific text on websites that the site administrator defines. If any of those web pages contain hyperlinks, it spiders through those pages, identifying the title, blurb, document type, and locate in a standard easy to read format. Some search engines are capable of searching Boolean expression, phrase, clause etc Lewandowski (2012).

STATEMENT OF PROBLEM

Nowadays, many internet users find it difficult to retrieve vital information from the Internet. Several performance schemes had been employed in order to evaluate the retrieval effectiveness of search engines. Yet user may

M. S. Bute, Department of Mathematics, Gombe State University, Gombe State, Nigeria.

Adamu Hussaini, Department of Mathematics, Gombe State University, Gombe State, Nigeria.

Inuwa Hadiza Adamu, Department of Mathematics, Gombe State University, Gombe State, Nigeria.

not know which search engine is best depending on information they require. Several types of search engines have been designed and implemented based on different retrieval methods, algorithms, and database technique. Hence, majority of the users finds it difficult to know which search engine is best. In this paper, the search engines performance will be evaluated based on retrieval effectiveness using three metrics; precision, recall, and response time.

SCOPE AND LIMITATION

In this research work, we are going to estimate the precision, response time and relative recall of five search engines Google, Yahoo, Bing, Baidu, and AOL. 24 search queries were sampled and these queries were categorized based on; Information, Navigational, and Transactional queries. This work is limited to only the first fifty (50) sites displayed by each search engine.

RELATED WORKS

Kumar and Bhadu (2013) this paper compare three Search Engines. The precision and relative recall of each search engine was considered for evaluating the performances of the search engines. Queries were tested, Results of the study showed that Google is the best amongst the search engines used.

Oberoi and Chopra (2009) finds that a web search engine opens the door to explore a huge amount of information. There is a variety of search engines which offer diversified services to its users. This paper draws a clear picture of the differences between various search engines and disproves the notion that all web search engines have same search capability, coverage, ranking and indexing techniques. Web search engines differ from each other in multiple aspects such as the searching strategy, coverage of the web, relevance of the search results with respect to the search query, ranking of the search results etcetera. The overlapping of the search results offered by the search engines is very low. The overlapping of the results from various search engines could be measured by collecting sample URLs from the result set of a search engine for a specific query. URLs from the collected data can then be matched with the results of another engine by performing a string comparison. The number of matches could be recorded to determine the fraction of URL overlap. Search engines with a single source have low web coverage in comparison to a meta-search engine.

METHODOLOGY

Information retrieval systems in general and specifically search engines need to be evaluated during the development processes as well as when the system is running in order to see how effective they are. The primary reason of the evaluations is to improve the quality of searching process and results, although there are other reasons for evaluating search engines. Measures used for search engines evaluation in this research work are:

PRECISION

This is the sum of the scores of sites retrieved by a search engine to the total number of site selected for evaluation Kumar and Bhadu. After the retrieval of search result, the user is sometime able to retrieve relevant information. The quality of searching the right information accurately would be the precision value of the search Kumar and Prakash (2009).

The search results which were retrieved by the search engines used were categorized as relevant, irrelevant, links and site can't be accessed.

If the web page is much close to the subject matter of the search query then it was categorized as relevant and two was given as the score. If the webpage is not related to the subject matter of the search query then it was categorized as irrelevant and zero was given as the score. If a web consists of a whole series of links, rather than the information required, then it was categorized as link and one was given as the score. If a message appears, "site can't be accessed" for a particular website the page it was categorized as site can't be access and zero was given as the score.

RESPONSE TIME

This is the time taken for each query search to be completed by the engines; it can be measured using stop clock or as is displayed by some search engines Kausar, Dhaka and Singh (2013)

RELATIVE RECALL

This is the ability of a system to retrieve all or most of the relevant documents in the collection. The relative recall can be calculated using the following formulae;

$$\text{Relative recall} = \frac{\text{total No.of sites retrieved by the search engine}}{\text{sum of sites retrieved by all the search engines}}$$

SEARCH ENGINES

Search engines are programs that search documents on the World Wide Web as requested by the user seeking for information Lewandowski (2008). Five search engines; Google, Yahoo, Bing, AOL, and Baidu were be used to determine the precision, response time and relative call of some sample search queries created for this research. Since more sites would be retrieved from the search engines for each search query, it was decided to select only the first 50 sites displayed by each of the search engines listed above for the evaluation Lewandowski (2013).

SEARCH QUERY

Search query is what a user inputs into a search engine and expects results as information Lewandowski (2009), Al-akashi and Inkpen (2012). A total of 24 queries were used for the evaluation, which is based on single word, double word, sentence, and alphanumeric queries and it is categorized into Informational, Navigational and Transactional queries in order to have an optimal results at the end of the research Kumar, Suri and Chauham (2005).

Table 1: Search Queries

SEARCH QUERIES USED FOR EVALUATION

S/N	SINGLE WORD QUERIES	DOUBLE WORD QUERIES	SENTENCE QUERIES	ALPHANUMERIC QUERIES
1	Q1.1 Aljazeera	Q2.1 NITDA Nigeria	Q3.1 Array data structure	Q4.1 HP laser jet p2035
2	Q1.2 Amazon	Q2.2 Skye Bank	Q3.2 Software prototype model	Q4.2 Visual Basic Programming Language 2010
3	Q1.3 Pseudocode	Q2.3 Symmetric Key	Q3.3 Artificial neural network	Q4.3 OS Android 5.0
4	Q1.4 Cryptography	Q2.4 Numerical Analysis	Q3.4 Integer division in C	Q4.4 Samsung Galaxy S3
5	Q1.5 Iphone	Q2.5 Dell Inspiron	Q3.5 Operating systems ebooks	Q4.5 Honda civic model 1998
6	Q1.6 Windows	Q2.6 Strong Decoder	Q3.6 Application of ant colony algorithm	Q4.6 2*2 matrix

Key**SQ: Search Query****RT: Response Time (sec)****SR: Search Results****NFSE: Number of Sites Used for Evaluation****R: Relevant****IR: Irrelevant****L: Link****CBS: Can't Be Accessed****P: Precision****Table 7: Overall Average Precision of Each Search Engine**

QUERIES	SEARCH ENGINES				
	Google	Yahoo	Bing	AOL	Baidu
Single Word Queries	1.48	1.26	1.31	1.31	1.35
Double Word Queries	1.28	1.51	1.33	1.32	1.32
Sentence Queries	1.38	1.34	1.41	1.16	1.38
Alphanumeric Queries	1.33	1.19	1.33	1.29	1.30
TOTAL	5.47	5.30	5.38	5.08	5.35
AVERAGE PRECISION	1.37	1.33	1.35	1.27	1.34

The sample queries were tested using the selected search engines and readings were taken as seen above from Table 2 to Table 6. Table 7 above shows the overall average precision of each search engines as it was computed from tables presented above. The study show Google has the highest precision against the other search engines.

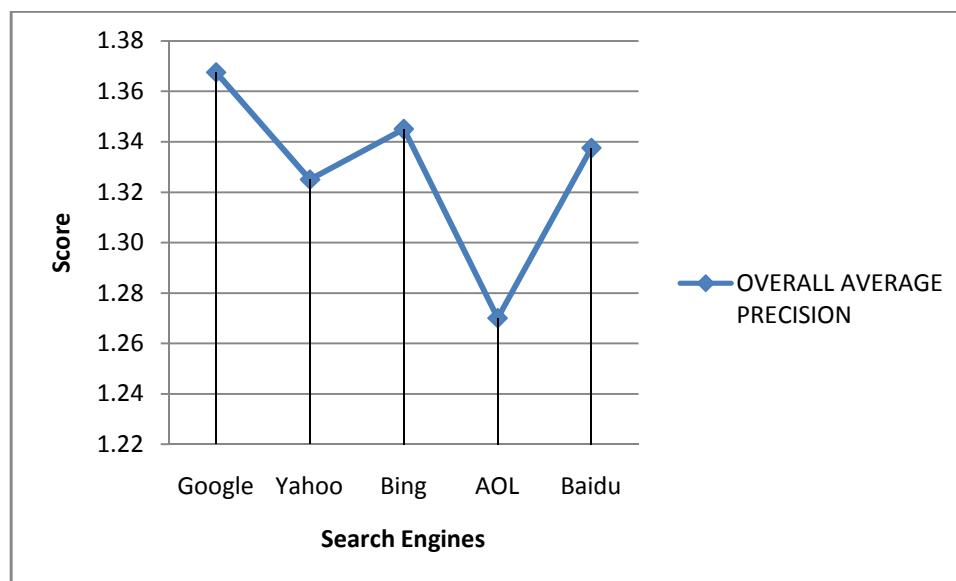
**Figure 1: Overall Average Precision of all Search Engines**

Table 8: Overall Average Response Time of Each Search Engine

QUERIES	SEARCH ENGINES				
	Google	Yahoo	Bing	AOL	Baidu
Single Word Queries	0.32	0.52	0.58	0.62	0.61
Double Word Queries	0.39	0.41	0.54	0.56	0.61
Sentence Queries	0.44	0.58	0.43	0.44	0.64
Alphanumeric Queries	0.40	0.55	0.57	0.61	0.73
TOTAL	1.55	2.06	2.12	2.23	2.59
AVERAGE RESPONSE TIME (sec)	0.39	0.52	0.53	0.56	0.65

From Table 8 above it shows that Google has the least response time, by that means Google is the fastest in terms of information retrieval.

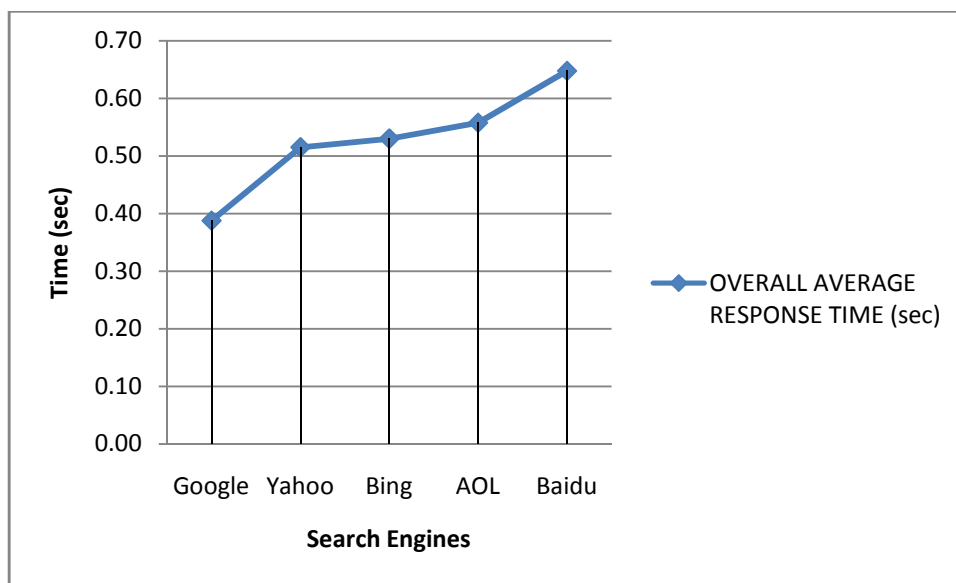
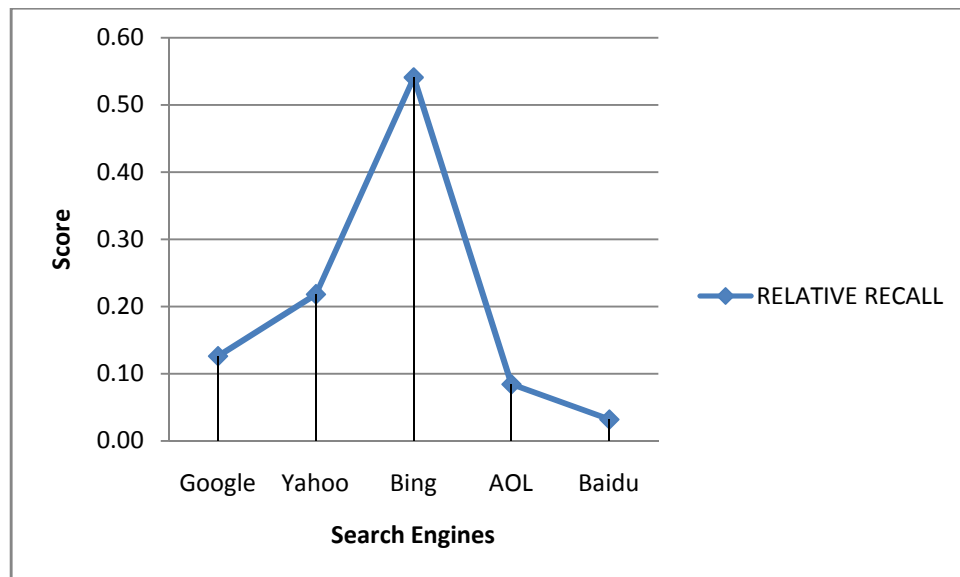


Figure 2: Overall Average Response Time of all Search Engines

Table 9: Relative Recall of Each Search Engine

QUERIES	SEARCH ENGINES				
	Google	Yahoo	Bing	AOL	Baidu
Single Word Queries	1,943,412,000	1,362,720,000	1,362,730,000	1,362,720,000	437,001,000
Double Word Queries	91,790,012	34,523,000	33,523,000	34,523,000	34,523,000
Sentence Queries	15,803,000	64,240,000	8,010,270,000	41,855,000	23,636,000
Alphanumeric Queries	143,744,000	2,342,285,408	30,688,000	23,578,000	55,683,200
TOTAL	2,194,749,012	3,803,768,408	9,437,211,000	1,462,676,000	550,843,200
RELATIVE RECALL	0.13	0.22	0.54	0.08	0.03

From table 9 above Bing has the highest relative recall, as it is the search engine that retrieves more search result than the other search engines.



CONCLUSION

This research paper compares the retrieval effectiveness of some search engines based on; precision, relative recall, and response time. The analysis was carried out within a specific period of time, it was noted that type of query affects the search effectiveness of search engines it was found that Google is best for single word queries, double word queries and alphanumeric queries while Bing is the best for sentence queries.

REFERENCES

- Akinola, O. S., 2010. A comparative study of some popular web search engines. *African journal of computer and ICT*, 3-20.
- Al-akashi, F. H and Inkpen, D., 2012. Query-Structure Based Web Page Indexing.
- Can, F., Nuray, R and Sevdik, A. B., 2004. Automatic performance evaluation of Web search engines. *Information Processing and Management*, 495-514.
- Kausar, M. A., Dhaka, V. S and Singh, S. K., 2013. Query-Structure webpage indexing Volume 63-No. *International Journal of Computer Application*.
- Kumar, K and Bhadu, V., 2013. A comparative study of BYG search engines. *American Journal of engineering research (AJER)*, 39-43.
- Kumar, R., Suri, P. K and Chauhan, R. K., 2005. Search Engines Evaluation. *DESIDOC Bulletin of Information Technology*, 3-10.
- Kumar, S and Prakash, N., 2009. Precision and relative recall of search engines: A comparative study of google and yahoo
- Lewandowski, D., 2009. The retrieval effectiveness of search engines on navigational. *ASLIB Proceedings*, (pp. 60-65).
- Lewandowski, D., 2012. A framework for evaluating the retrieval effectiveness of search engines. *Journal of the American society for Information science and Technology*, 20-30.
- Lowandowski, D., 2008. The Retrieval effectiveness of Web search engines: Considering results description. *Journal of documentation*
- Lowandowski, D., 2013. Evaluating the retrieval effectiveness of Web search engines using representative query sample. *Journal of the American for Information science and Technology*.
- Oberoi, S. O and Chopra, M., 2009. Web Search Engines- A Comparative Study.

