

COMPARATIVE ANALYSIS OF SEARCH ENGINES ON INFORMATION RETRIEVAL BY UNDERGRADUATE STUDENTS: (A CASE STUDY OF FEDERAL UNIVERSITY OF TECHNOLOGY MINNA, NIGER STATE)

SAMUEL J. UDOUDOH, PHD

*Department of Library and Information Technology,
Federal University of Technology, Minna, Niger State.*

Abstract

The study was designed to examine the Comparative Analysis of Search Engines Utilization among Undergraduates, using the students of the Federal University of Technology, Minna as a Case Study. Three research questions were raised and answered. The instrument used for data collection was validated research-made questionnaire christened: Undergraduate Students Comparative Search Engines in Usage Analysis Inventory Questionnaire (USCSEUIQ), with a reliability index of 0.68. A sample size of 200 respondents was the population of the study, with a set of 40 students drawn randomly from 5 departments; Electrical/Electronics Engineering, Library & Information Technology Mathematics/Computer, Microbiology and the Physics/Electronics. The two campuses of Bosso and Gidan Kwano respectively participated. The simple percentage analysis was used to analyse the data generated by this study. The analysis revealed that Web search technology has been evolving very rapidly and will continue to evolve. The analysis revealed that students tend to have more confidence in their search engine of choice than in their own ability to formulate a query that will return their desired results. This study reveals that Google is remarkably better than Yahoo and MSN search engine and the study further confirmed that Google surpasses all other search engines and is the most widely used presently. The study concludes that speed, relevance and effectiveness determine which internet search engine to use.

Background of the Study

A search engine is a computer program which acts as a tool or means of retrieving information from a database, based on certain criteria defined by the user. Modern search engines search databases which contain large amount of data, accessed from the World Wide Web (w.w.w.), news group and directory projects. The evolution of the search engine was rather quick, with the earliest true search engine appearing at the beginning of the 1990s, and the first modern style search engine appearing in 1995 (Brendan, 2006). Winship (1995) in his opinion confirmed that, search engines have become the most important tool for information seeking and retrieval. Based on their tremendous economic value, search engine companies constantly put major efforts to improve their search results. The measurement of search effectiveness is thus an important issue. Although many evaluations have been done on different search engines in the past, they mainly use fixed sets of queries and judge the relevance of each returned page by a panel of human judges (*Bing 2006*). The results were often measured on precision and recall just like information retrieval. However, this evaluation method was by no means ideal because relevance did not mean user satisfaction can only be measured using queries from the user's daily information needs and based on his/her personal assessment of utility of the returned results to the queries can ideal evaluation is a personal evaluation according to (Bar-Ilan, 2004).

At the end of 2005, Google was the search engine with the highest number of users in the world, including a particularly high market share in France with (82%) of traffic according to Xitil. The reasons why a user might choose one search engine over another are complex, while elements such as speed, ergonomics and aesthetics all come into play, the most important criterion seems to be the relevance of the results to the search performed, at least in the way they are perceived as relevant by the user, (Veronis, 2006). Secondly, we are a society obsessed with convenience. We go to extreme lengths to invent devices that promise a simpler or more convenient lifestyle. This paradox is exemplified in our fascination with the internet, as well as with our attempts to index it for information access purposes. The internet is today a rapidly evolving organism that is almost completely lacking in fundamental organization. The question of whether each individual achieves a net gain from all the efforts expended in this process lies somewhere beyond the scope of this study but the researcher can agree on the need to properly organize this much unstructured information

resource.

Internet search tools have been created to answer this very pressing need. They are evolving rapidly some would say more rapidly than the internet itself. By the end of 1996, it was estimated that the internet will consist of no less than 150 million pages, containing 50 or 60 billion words. To make matters worse, this great mass of data exists completely without any kind of bibliographical controls, standard numbering systems, or classification systems. Clearly, automated tools of some sort are necessary to sift through this mass of material (Venditto, 1996). This study was designed to compare and analyze the effectiveness and efficiency of different search engines and information retrieval used by final year students for their research studies. This research was also to test the relevance of the search results to the students search queries.

Statement of the Problem

The use of search engines and their advanced functions is often fraught with difficulty. There are a lot of problems associated with the use of search engines, for example: web searcher not always finding what they are searching for online, i.e. high proportion of irrelevant documents may be retrieved. There is also a problem of efficiency and effectiveness of the search tools engines. However, due to the information redundancy on the web, a researcher easily finds a huge number of relevant pages to almost any query. Thus page relevance is no longer a major issue. The usefulness/utility of each top ranked page to each individual user becomes the key. The evaluation of usefulness can only be done based on queries derived from the user's personal information needs and his/her personal perception of the returned results to the queries. It was this belief that guided this evaluation. Such an evaluation can truly tell us why users choose/or prefer one search engine over another.

Apart from these problems, there are too many electronic search engines available to track by hand. The massive growth of the search engines is parallel to the growth in information searching and retrieval becoming available on the deep web. Search engines often contain dynamic data, which constantly grows and changes. Another problem is that many search engines are not index able by traditional methods. Only surface web documents have a hyper linking system that supports easy indexing (Berkeley Digital Library SunSITE, 1996).

Objectives of the Study

The following were the objectives of this research investigation:

1. To analyze the frequency of usage of search engines among students.
2. To find out the most widely used search among the students.
3. To determine why students prefer one search engine over the other.

Research Questions

The following research questions were to guide this study:

- (i) How frequent do students use search engines for research?
- (ii) Which search engine do students use most?
- (iii) How does a search engine affect students' research work?

Scope of Study

This research work was confined to the final year students of the Federal University of Technology, Minna embodying students in two schools; School of Science and School of Engineering respectively while the evaluation was limited to five different departments within these two schools.

Literature Review

According to *Webopedia* (2008), a search engine is defined as a program that searches documents for specified keywords and returns a list of the documents where the keywords were found. A search engine is really a general class of programs; the term *search engine* is used in this study to specifically describe systems like Google, Alta Vista and Excite which enable users to search for documents on the World Wide Web and USENET newsgroups. Typically, a search engine works by sending out a *spider* to fetch as many documents as possible. Another program called an *indexer* then reads these documents and creates an index based on the words contained in each document, each search engine uses a propriety algorithm to create its indices such that ideally, only meaningful results are returned for each query.

Review of Previous Study on Search Engines Usage

Verona'sd, (2006) used students to evaluate 14 topic areas using six search engines and his studies showed that Google and Yahoo were the best engines although they did achieve the pass grade of "2.5". Lewandowski (2008) conducted a survey on 5 major search engines

(Google, Yahoo!, MSN, Seekport and ASK) and discovered Google was ranked higher and preferred by students. However, users are adopting different information seeking strategies than those used in more traditional contexts (Jansen, Spink and Saracevic (2000). This has also been recorded by Marchionini (1992), who stated that "humans will seek the path of least cognitive resistance" and Griffiths (1996), who found that "increasing the cognitive burden placed on the user... can affect successful retrieval of information. Where an application required fewer actions from the user, greater success was achieved as there was less possibility for a user to make an error".

An informative review of Web searching studies by Jansen and Pooch (2001) compares the searching characteristics of Web information seekers with those of users of traditional IR systems, but their study separates out online public access catalogue (OPAC) users from general IR system users. So, for example, they found that OPAC searchers express their information needs in queries of one to two terms, while Web searchers use approximately two terms and IR searchers six to nine terms per query. Searching session length also differed, with web searchers usually using two queries per session and typically viewing no more than ten documents from the results list, OPAC searchers using two to five queries and viewing ten documents per session. In addition, while 37 percent of IR searchers use Boolean operators, only 8 percent of web searchers and 1 percent of OPAC searchers use more advanced searches.

Other observations of the average web researcher (Spink, Wilson, Ellis and Ford, 1998; Ellis Ford, and Former, 1998) point out that in effective use may be caused by a lack of understanding of how a search engine interprets a query. Few users are aware of whether or not a search service defaults to "and" or "or" and expect a search engine to automatically discriminate between single terms and phrases. Also, devices such as relevance feedback work well if the user ranks ten or more items, when in reality users will only rank one or two items for feedback Croft, (1995). Koll (1993) found that users provide few clues as to what they want, approaching a search with an attitude of "I'll know it when I see it, "which creates difficulties in formulation of a query statement.

Larsen (1997) is of the opinion that internet search systems will evolve to meet the behavior of the average web searcher. Thus it can be seen that there has been a shift toward the introduction of search features that appear to respond to the ways in which users actually search these systems, for

example, search assistance, query formulation, query modification, and navigation. The notion that improved interaction is key to improving results is attractive in principle but not necessarily true in reality. Nick Lethaby of Verity incorporated, paraphrased in Andrews (1996), pointed out that users do not want to interact with a system beyond entering in a few keywords.

A separate research project conducted to develop a methodology for the evaluation of internet search engines from a user's perspective (DEVISE - Dimensions in Evaluation of Internet Search Engines) conducted to develop a methodology for the evaluation of internet search engines from a user • perspective also found that interaction had little value by users as the interaction dimensions and the weakest correlation with user's overall rating of satisfaction, where efficiency had the strongest correlation, followed by Effectiveness, Utility, and then Interaction (Johnson, Griffiths, and Hartley, (2003). It can thus be assumed that most users will not use advanced search features, nor enter complex queries, nor want to interact with search systems. As a consequence, systems such as search engines are now trying to automate query formulation, shifting the burden of formulating precise or extensive terminology from the user to the system.

Mode of Operations of Search Engines

The search engines do not really search the World Wide Web directly. Each one searches a database of web pages that it has harvested and cached (Tilman, 2003). When a search engine is used, you are always searching a somewhat state copy of the real web page. When the searcher clicks on links provided in a search engine's search results, he retrieves the current version of the page. Search engine data bases are selected and built by computer robot programs called spiders. These "crawl" the web, finding pages for potential inclusion by following the links in the pages they already have in their database. They cannot use imagination or enter terms in search boxes that they find on web. If a web page is never linked from any other page, search engine spiders cannot find it. The only way a brand new page can get into a search engine is for other pages to link to it, or for a human to submit its URL for inclusion. All major search engines offer ways to do this.

After spiders find pages, they pass them on to another computer program for "indexing". This program identifies the text, links, and other content in the page and stores it in the search engine database's files so that the database can be searched by keyword and whatever more

advanced approaches are offered, and the page will be found if your search matches its content. Many web pages are excluded from most search engines by policy as the contents of most of the searchable databases mounted on the web, such as library catalogs and article databases, are excluded because search engine spiders cannot access them. All this material is referred to as the "Invisible Web" what you don see in search engine results (Tilman, 2003).

Measures of Effectiveness and Efficiency of Search engines

A document is said to be relevant if it is the one which the user(s) perceive as containing information of value with respect to their personal information search and needs. In other words, a relevant document is one that a person judges as useful in the context of his specific information needs. However, evaluation is the key to making progress in building better search engines. It is also essential to understanding if a search engine is being used effectively in a specific application. One of the primary distinctions made in the evaluation of search engines is between *effectiveness and efficiency*. Effectiveness, loosely speaking, measures the ability of the search engine to find the right information, and efficiency measures how quickly this is done. Effectiveness of a search engine is the quality of its search result. It is a measure of the ability of the system to satisfy the user in terms or relevance of documents retrieved. Meanwhile, efficiency is defined in terms of the time and space requirements for the algorithm that produces the ranking (Bar-Han, 2004).

According to Cleverdon (1996), there are six measurable quantities of evaluating an information retrieval model or a search engine:

- ◆ The coverage of the collection, that is, the extent to which the system includes relevant matter.
- ◆ The time lag, that is, the average interval between the time the search request is made and the time an answer is given.
- ◆ The form of presentation of the output
- ◆ The effort involved on the part of the user in obtaining answers to his search requests
- ◆ The recall of the system, that is, the proportion of relevant material actually retrieved in answer to a search request.
- ◆ The precision of the system, that is, the proportion of retrieved material that is actually relevant.

Table 1: How long have you been an internet user?

Years of Experience	Frequency	Percentage (%)
Below 1 year	46	23
2-5 years	74	37
Above 5 years	80	40
Total	200	100

Table 1 above shows user year (s) of internet use experience range between year 1 to 5 years respectively. The table reveals that only 80(40%) of users have more than 5 years of internet use experience, 74(37%) had between two to five years experience while 46(23%) less than one year of experience.

Table 2: Analysis of internet search engines usage frequency over 5 years period

S/ N.	Search Engine	Very Often	Percent- age (%)	Often	Percent- age (%)	Rarely	Percent- age (%)	Not at all	Percent- age (%)
1	Yahoo	30	20	29	19.5	39	19.5	40	11
2	Google	80	43	48	16	5	2.5	0	0
3	MSN	34	12	14	12	40	20	52	35
4	Alta Vista	7	3.5	27	12	45	22.5	10	5
5	Lycos	4	2	12	6	16	8	0	0
6	AOL	7	3	15	7.5	11	5.5	15	7.5
7	Ask	22	8.5	31	15.5	15	7.5	36	18
8	Net- scape	6	3	17	8	28	14	10	5
9	Mamma	10	5	7	3.5	1	0.5	37	18.5
	Total	200	100	200	100	200	100	200	100

From table 2 above, the respondents using the search engine very often over a five year period; Google was rated 43% (80) ahead of Yahoo search 20% (30), MSN was rated 12% (34), while, 7.5%, 18%, 5% and 18.5% of the students do not at all use AOL, Ask, and Netscape, respectively.

Table 3: Rating of Search Engines Effectiveness and Relevance

S/N.	Search Engine	Very Effective	%	Effective	%	Not Effective	%
1	Yahoo	32	5.5	28	14	45	22.5
2	Google	75	55	60	30	2	1
3	MSN	16	8	42	21	52	26
4	Alta Vista	12	4	8	4	14	7
5	Lycos	10	5	8	4	5	2.5
6	AOL	9	4.5	42	21	10	5
7	Ask	24	12	4	2	25	12.5
8	Netscape	7	3.5	5	2.5	10	5
9	Mamma	15	2.5	3	1.5	37	18.5
	Total	200	100	200	100	200	100

Table 3: above indicate that Google search engine is rated the best in terms of effectiveness and relevance of search results with 55% (75) of the respondents who rated Google search as very relevant, ahead of Yahoo search 5.5%(32), (105) of the respondents. Over 40% said believed that other engines were not relevant in terms of their search results, i.e. Alta Vista, Lycos, AOL, Ask etc.

Table 4: Usefulness and relevance of search engines in research work**a) Search Engine usefulness in research work**

Rating Options	Frequency	(%)
Strongly Agree	152	76
Agree	43	21.5
Disagree	2	1.0
Strongly disagree	3	1.5
Total	200	100.0

b) Extent of getting relevant results from search engine

Rating Options	Frequency	(%)
Strongly Agree	90	45.0
Agree	77	38.5
Disagree	27	13.5
Strongly disagree	6	3.0
Total	200	100.0

Table 4a: above indicates that 76 % (152) of the respondents strongly agree that search engine was very helpful in their research work, 21.5% (43) Agree, 1 % (2) disagree and slightly above 1.5 % (3) of the students strongly disagree with the fact. Table 4b: above indicated that 45% (90) respondents strongly agree that they always get relevant results from internet search engines, 38.5% (77) also agree to this, while 13.5% (27) disagree while only 3%(6) respondents strongly disagree to this fact.

Discussion of Results

The data analyzed above sought to measure students' perceptions of the Internet usage or search engines for their research works.

How frequent do students use search engine for research?

This question sought to know the level of student awareness and frequency of usage of the various search engines. The subject of study revealed that students possess great awareness of search engines with 43% (80) of them having had more than 5 years of internet search engines experience. This showed that almost all the students sampled were relatively aware of internet search engine and its usage for information retrieval.

Which search engine do students use most?

The research question above was asked to know the most widely used search engines among the students. From the result, Google was identified as the most widely searched engine. Over 80% of the participants rated Google as their most favorite while 50% preferred Yahoo search engine. The study also indicated that 17% of respondents preferred MSN search, 13.5%, 11% 3.5% choose ASK, Mamma, and AOL respectively.

How does a search engine affect student's research work?

The above research question was asked to know the general information about how helpful was the search engine to students' research activities. 76% of the respondents strongly agree that search engine was very helpful in their research work, 21.5% agree, 1% disagrees with the fact.

Summary

The researcher does not expect most of the findings in this study as a surprise to most search marketer too. What may be enlightening for example is the clear domination of Google as the most preferred search engine. However, the researcher believes that the results truly reflect the users sentiments based on their personal information needs and

assessments demand. It is interesting to observe that Google has one of the largest databases of Web pages including many other types of web documents: (blog posts, wiki pages, group discussion, threads and document formats (e.g. *pdf, Word or Excel documents, Power Point*)).

Furthermore, it was discovered that students who prefer Google, only use Google and never another one. They usually use the same one or several search engines for all their choice of searches regardless of the type of information being sought. The analysis revealed that students tend to have more confidence in their search engine of choice than in their own ability to formulate a query that will return their desired results choosing to re-launch a modified version of their original query in the same engine before giving up on that engine's ability to return satisfactory result.

Conclusion

Web search technology has been evolving very rapidly and will continue to evolve. This study reveals that Google is remarkably better than Yahoo and MSN search engine. However, Google may be reaching the limit of the current search paradigm. Further improvement by Google will take much more effort (possibly exponential amount of effort for linear gain). It is thus time for Yahoo!, MSN search and others to catch up, which is easier to do, if they are ready. Based on the evaluation results, the researcher believes that it is going to be very hard for either one of them to overtake Google (unless Google makes hard decisions), but to get close to Google is very unlikely, which in my opinion will happen in a not-distant future. There are good search engines, but there is no perfect search engine. This is because internet is subjective to the end user, making it impossible for any search engine to understand every person's intent correctly each and every time. Furthermore, it still seems hard to evaluate which search engine is the best because it is really just a matter of opinion.

However this research ascertained that Google surpasses all other search engines and Google is the most widely used among students because of its effectiveness, efficiency and relevancy of the search results to their search queries and information needs.

Recommendations

Based on the findings of this study, the following recommendations are made: to help students in their choice of search engines and in formulating a better search queries that will give them the desired and

satisfactory results:

- (i) Students should have more confidence in their own ability to formulate a query that will obtain a needed result than in the search engine.
- (ii) Students should endeavor to use several search engines for a particular search query and compare the results than using the same search engine over and over.
- (iii) Internet users generally should acquaint themselves with the advanced search features on several engines for better query formulation.
- (iv) There should be frequent update of databases of web pages so that outdated information would be replaced with recent information
- (v) Finally, the study reveals that majority of the students sampled strongly believe that speed, relevancy, and effectiveness determine which search engines one would choose to use. The study concluded that search engines designers and marketers should improve in these areas for better performance and if they are to be relevant in this field.

REFERENCES

- Aina, L. O. (2002). *Research in Informational Sciences: AN African Perspective* Ibadan Stirling - Hordon P. 19. Andrews, W. (1996), *Search engines gain tools for sifting content on the fly*. New York
- Bar-Ilan, J. (2004). The use of web search engines information science research. In B. Cronin (Ed.) *Annual Review of information science and technology* (Vol. 33, pp. 231-288) Medford, Ny, USA:
- Bing Liu, (2006). *Personal Evaluation of search engines*, Chicago, Retrieved from: <http://sunsite.berkeley.edu/Help/searchdetails.html>.
- Brendan, M. (2006). *Conjecture corporation*, Retrieved from: <http://www.wisegeek.com/what-was-the-first-search-engine.htm>.
- Chang, N. C. & Perng, J.H. (2001). Information search habits of graduate students at Tatung University. *International information & Library Review*, 33 (4), pp. 341-346
- Edwards, S. L., & Bruce, C.S. (2002). Reflective internet searching: an action research model. *The learning organization: An International Journal*, 9 (4).

pp. 180-188.

- Jansen, B.J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: a study and analysis of user queries on the web, *information processing & management*, 36 (2), pp. 207-227
- Jansen, B. J. and Pooch U. (2000). A review of web searching studies and a framework for future research: *Journal of the American Society for Information Science* 52 (3): pp 236-46.
- Jilian, R. G. (2009). "Student searching behaviour and the web: use of academic resources and Google". Library trends. Find Articles.com, retrieved from http://fmdarticles.com/p/articles/mi_m1387/is_453/ai_nl4732768/
- Johnson, F., Griffiths J.R. and Hertley, R.J. (2003). Task dimensions of user evaluations of information retrieval systems. *Information research*, 8 (4) pp 157, retrieved 15 July 2003 from <http://nformaitonR.net/ir/paper157.html>.
- Lewandowski, D (2008). The retrieval effective of web search engines: a considering results description.
- Spink, A., Bateman, J., & Jansen, B. J. (1999). Searching the Web: A survey of EXCITE users. *Internet Research: Electronic Networking Applications and Policy*, 9 (2), 117-128.
- Spink, A., & Jansen, B. (2004). A study of Web Search Trends; *Webopology*, 1(2). Article 4, retrieved from <http://www.webology.ir/2004/v/ln2/a4.html>.
- Sullivan, D. (2000). The Webmaster's guide to search engines and directories, retrieved from: <http://calafia.com/webmasters/study.html>
- Tillman, H.N. (2003). Evaluating quality on the Net; Retrieved from, <http://www.hoptillman.com/findqual.html>
- Veronis, J. (2006). The comparative study of six research engines; Retrieved from, <http://www.up.univ-mrs.fr/veronis>
- Winship, Ian (1995). "World wide web searching tools." *Vine* 99 :pp. 49-54.
- Zipf, G. (1949). *Human Behaviour and the Principle of least effort*, Reading, M. A: Addison Wesley http://en.wikipedia.org/wiki/search_engine