

PERFORMANCE OF TEN INTERNET SEARCH ENGINES IN RETRIEVING SCIENTIFIC LITERATURE IN TANZANIAN PUBLIC UNIVERSITY LIBRARIES

L.S.P. Busagala¹ and T. Elly²

Abstract

Several public Search Engines exist of which their coverage and response time differ depending on various factors including connection speed. Now, which one does perform best under the present connection speeds? This study aimed at investigating the information indexing and retrieval effectiveness and efficiency of ten selected search engines under different connection speeds at two University Libraries namely Sokoine National Agricultural Library (SNAL) at Sokoine University of Agriculture and University of Dar es salaam Library (UDSM). Google followed by Yahoo outperformed all the other eight-search engines in terms of relevance, precision, and responsiveness. In terms of other criteria such as phrase searching, simple and natural language interface, high quality of display results, these search engines were the best. MetaSearch engines especially MetaCrawler performed the worst in indexing and retrieving scientific literature particularly at UDSM library. There was a significant difference of search engines performances between the two connection speeds.

1. Introduction

1.1 Background information

The Internet technology has revolutionized the way information can be acquired, processed, used and disseminated. Scientific literatures are reported to be disorganised on the internet and Information users are overwhelmed by a lot of information most of which is irrelevant (Lawrence et al 1998, 2001). Searching for literature on the Internet may be time consuming and complex process (Lesser 2000; Lawrence, 2001; Sasikala and Patnaik, 1999; Hanka and Fuka 2000). Vidmar (1999) claims that understanding how search tools work, selecting a search tool to use for which purposes, identifying which tool does what best, is an unpleasant exercise even to the most dedicated cyber searchers.

¹ Busagala is Assistant Librarian, Sokoine University of Agriculture

² Elly is Assistant Lecturer, Faculty of Commerce and Management, University of Dar es Salaam.

Search engines simplify the exercise of locating and retrieving specific scientific literature on the internet (Niuwenhuysen, 1999, Proctor, 1997). A Web search engine is a computer program that searches for documents containing key words or phrases of interest to users. These are software programs called robots or accessible spiders that crawl through the files on the Internet and download primarily the gathered information (Lawrence et al, 2001).

Free or relatively cheaper scientific literature might be available on the internet (De Smet and Kertens, 2000). Bar-Ilan (2000) did a content analysis. He analyzed the Web pages retrieved by the major search engines. The list of references obtained from the Web was compared to data retrieved from commercial databases. For most cases, the list of references extracted from the Web outperformed the commercial, bibliographic databases. The results of these comparisons indicate that valuable, freely available data is hidden in the Web waiting to be extracted from the millions of Web pages.

However, Kibirige and De Palo (2000) reported that the coverage of search engines is less than a half and search results may be misleading. For example Excite uses concept-based clustering and Infoseek uses morphology concept such that they may return documents with unrelated words. Introna and Nissenbaum (2000) urged that research results suggest that search engines systematically exclude, in some cases by design and some accidentally certain web sites and index some web sites in favour of others, systematically giving prominence to some at the expense of others. Exclusion differs from one search engine to another leading to the need for evaluating their performance and to present the results to the information professionals and users.

Moreover, end user searchers may search infrequently and usually have fewer information skills than their professionals in information Technology do. A new range of databases and search engines with natural language interfaces are required to assist end users (Weimer and Rusch 1996; Tomaiuolo and Packer 1996). Simple interfaces and natural language searching that respond to the needs of the typical searcher are required (Vidmar, 1999). Therefore evaluating useful search features was included in the study.

Since search engines vary their coverage and performance (Proctor, 1997), which one does perform best under the present connection speeds? This study aimed at evaluating ten selected Search Engines in retrieving scientific literature.

1.2 Problem statement

Among the main problems in Tanzanian libraries is the lack of adequate funds (Tweve, 2000; Wema, 2000; Mnyani, 2000). Use of the Internet in identifying and utilising the available information for academic and research

is relatively cheaper as considerable amount of free articles are available online (De Smet and Kertens 2000). If a cost is involved it is usually lower than the cost of traditional means of acquiring information (Proctor 1997 and Levey 2001). Lawrence (2001) explains that the Internet provides convenient access to the increasing amount of literature and maximising the usage of scientific records benefits the society at large. He further argues that although availability varies greatly by discipline, over a million research articles are freely available on the web. Kibirige and De Palo (2000) state that search engines provide the most common access points utilised by library/information centre users to get to electronic resources on the Internet. Several public search engines and meta-systems exist.

However, the coverage of any one engine is significantly limited: no single engine indexes more than about one third of the indictable web (Lawrence and Giles 1998 and 1999). Also the coverage of search engines differs. Often, they do return documents that do not contain the query terms. Furthermore little information on the performance of search engines is available in Tanzanian University Libraries. Using all search engines at a time may be clumsy and difficult. Selecting one or more without having its performance records may be time consuming (Proctor 1997). This study therefore investigated the information retrieval effectiveness and efficiency of ten selected search engines under different connection speeds at two University Libraries namely Sokoine University of Agriculture and University of Dar es salaam.

1.3 Objectives of the Study

1.3.1 General Objective

To identify the effective and efficient search engines in retrieving scientific literature so as to recommend to academicians, researchers and other information users such that they can significantly improve their information searching behaviour by appropriately selecting search engines to meet their information needs.

1.3.2 Specific Objectives

- To identify the most effective search engines based on the relevance and precision of the retrieved information.
- To identify the most efficient search engines basing on the response time under different speeds of connectivity.
- To identify search engines that support phrase query searching as compared to other useful search features.

1.3.3 (a) Research Questions

- Which search engines are the most effective basing on the relevance and precision of the retrieved information?
- Which search engines are most efficient in information basing on the response time under the present connectivity?
- Which search engines that support phrase query searching as compared to other useful search features?

(b) Hypothesis

- Often, search engines do return documents that do not contain the query terms hence irrelevant literature and low precision.
- Most of the search engines are slow, respond after waiting for a considerable time.
- Most search engines do not support phrase query terms or natural language of the user and they have limited available useful search features.

1.3.4 Significance of the Study

As mentioned before, Tanzanian University Libraries face financial constraints in acquiring reference materials. Furthermore, outdated information is always available in their collections (Wema, 2000; Mnyani, 2000 and Katundu, 1998). The Internet technologies are relatively cheaper in acquiring information that would not be acquired through different methods (Proctor 1997 and Levey 2001). Although it is relatively cheaper to acquire literature on the internet, it is however worthless using considerably a lot of money for internet subscriptions without effective and efficient use of the internet. For instance, about US \$3,000 per month at Sokoine University of Agriculture was used for Internet subscription, which was not effectively utilised as users had limited information on the effectiveness and efficiency of search engines under the available connection speeds.

Since there are several search engines the study is helpful to academicians, researchers, students and other information searchers to make them appropriately select search engines to meet their information need.

The findings also have a role of increasing the literature base and information access for various purposes such as for research and other developmental programmes because evidence shows that information usage increases when access is more convenient and maximising the usage of scientific literature benefits the whole society (Lawrence 2001).

2. Methodology

2.1 Place for experiments

Experimental tests were used in this study. Sokoine University of Agriculture Library, namely Sokoine National Agricultural Library (SNAL) and University of Dar es Salaam (UDSM) library were chosen as places for the experiments. These libraries are connected to the Internet. SNAL shares its connection speed with the entire University, which shares 1Mbps with several other universities. University of Dar es Salaam library also shares its 1Mbps connection speed with other campuses that include Muhimbili College of Health Sciences, University College of Lands and Architectural Studies and Hubert Kariuki University.

2.2 Constraints and assumptions put on the experimental tests

Internet connection speed depends on various factors including the time of day. In order to get high quality findings the following constraints and assumptions were put into consideration: -

Random errors – errors that are irregular with respect to time and hence cannot be predicted. These were minimized by recording five values of the same query (index and retrieval). Finally taking the average (mean) of all values (Chatfield 1983).

Counting search terms used in searches that were viewed in retrieved documents.

Five (5) queries were run on ten search engines and number of results returned by each engine was reported. This approach was important to minimise the error of making wrong inferences.

In query formulation more than one or two words is desirable for more relevant and precise results. A one word query with a very common term ends up with millions of results of which most are irrelevant (Becket 1998). Therefore a query had more than one word.

It was important to know the available bandwidth at each time during the experimental tests. This was important because search engines may be limited by the available network bandwidth (Lawrence 1998). For every one and half-hours the available bandwidth was examined by using the online services provided by search.com. The average bandwidth was 25.5 and 9.58 kbps at Sokoine National Agricultural Library and University of Dar es Salaam Library respectively.

The browser used was Netscape mainly version 4.5 to avoid errors that could be caused by the use of different browsers.

Since connection speed may vary in time of a day, experiments were done at similar time of day to avoid biases that might be due to congestion at peak hours. This was done with assumption that at similar time search engines may exhibit same or similar performance.

Assuming that Local Area Networks (LANs) were similarly set up and basing on this assumption experiments were conducted. Also it is useful to note that findings provide indicative information on the performance of searching tools on the internet. Changes may occur in the future due to the increased bandwidth and network set up changes.

2.3 Specification of Computers used in the experiments

The computer used at Sokoine National Agricultural Library had the following specifications: -

- Processor speed 366 MHz
- SDRAM 32MB
- Storage Disk Space 4.3GB

The computer used at the University of Dar es Salaam had the following specifications: -

- Processor Speed 366Mhz
- SDRAM 64MB
- Storage Disk Space 4.0GB

2.4 Selection of search engines to evaluate

It was important to select ten or less search engines due to the fact that time would not be enough to work with more search engines. Most researchers who probably worked with the search engines did their research at a different connection speed. Consideration was taken on the fact that the University libraries face financial constraints. Users always choose search tools that are accessible at no cost. Due to this fact it was reasonable to evaluate search engines that are available to users free of charge. Chu and Rosenthal (1996) argues that these free services might continue to be available to the Internet community in the foreseeable future.

2.5 Query Formulation

Five queries were selected basing on the information needs of five postgraduate students of Sokoine University of Agriculture. The queries consisted of the following words: -

- Query 1-> tephrosia vogelii maize growth yield
- Query 2-> gypsum bean yield salt soils
- Query 3-> wood ash rice straw nutrient ruminant
- Query 4 -> azolla caroliniana paddy yield
- Query 5 -> ethnoveterinary aloe

Since every search engine may require a specific syntax in order to retrieve relevant results consideration was taken in this regard.

2.6 Experiments

As mentioned before, the experiments were carried out at similar times of the day using the formulated queries. The syntax of the queries slightly varied depending on the recommendations of the search engine developer in their documentation. However phrase query experimentation was done to each search engines. For each query five tests were done. Counting of indexed and retrieved documents was performed.

Consideration of relevant documents was done up to the twentieth document because in the context of the World Wide Web, usually on the first ten or twenty documents are examined for relevance and computed for precision. This is due to the fact that web users hardly ever go beyond the first twenty hits and that, search engines rank their results according to relevance. The documents that have all and more words of the search terms are ranked highly (Bar-Ilan 2000; Chu and Rosenthal, 1996; Dong and Su, 1997; Salton, 1989)

2.7 Data Analysis

Data from experimental tests were summarized and computations of various quantitative measures were done. Values such as mean precision, mean response time and number of retrieved relevant documents were obtained. Furthermore statistical tests for significance of results were adopted. In the statistical tests t-distribution was adopted because of the following conditions: the sample size (n) was less than 25 or 30. The standard deviation (σ) was not known and the assumption that the population from which the sample is drawn was approximately normally distributed (Mann 1995, Chatfield 1983).

2.8 Measures for Evaluating Retrieved documents from the Internet

1) Relevance

There have been various ways used to evaluate relevance of the documents retrieved from information systems. Relevance may base on the user judgment. The user posing the query may judge the relevance of the document. This is the kind of judgment that may be considered as a subjective measure. Therefore this kind of measure was not used in this research.

Another way of evaluating relevance is through counting only the terms used in searches that are viewed more than once. This can be considered as technical relevance (Allen 2001). Technical relevance is measured by considering the retrieved document relevant if it contains all search terms or phrases. Furthermore such a document should miss all terms or phrases that are supposed to be missing that is terms preceded by minus sign or a NOT operator. Technical relevance is an objective and easily checkable measure

(Bar-Ilan 2000). In this study the technical relevance was used. One of the limitations of this measure is that a document may have all the criteria but on the part of the users it is irrelevant.

In the context of the World Wide Web, usually only the first ten of twenty documents are examined for relevance and computed for precision. This is due to the fact that web users hardly ever go beyond the first twenty hits and that search engines rank their results according to relevance. The documents that have all and more words of the search terms are ranked highly (Bar-Ilan 2000; Lawrence, 1998).

2) Precision

Precision p is another measure of information retrieval systems (Salton, 1989). It is a percentage of relevant retrieved documents out of the total number of documents retrieved by the system on a query. The expression below (1) shows how precision can be calculated and therefore was used in the evaluation of search engine performance.

$$p = \frac{r}{r + n} \times 100 \dots\dots\dots (1)$$

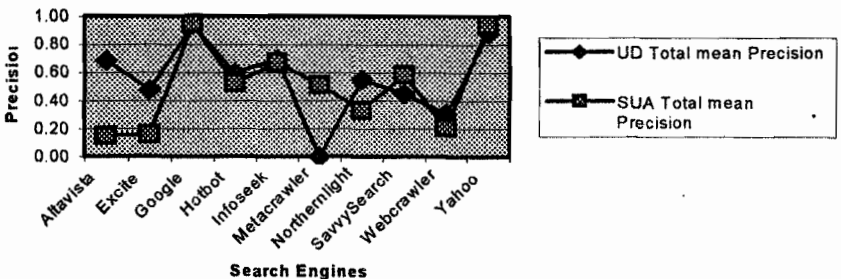
Where r is the number of relevant documents and n is the number of irrelevant documents. In the context of World Wide Web this measure might be limited because users usually never go beyond first twenty of retrieved documents (Bar-Ilan 2000; Lawrence, 1998).

3.0 Results and Discussion

3.1 Relevance and Precision of search engines

A document is considered technically relevant if it contains all search terms. Precision was obtained using formula (1) above. Table 1 shows the precision of each evaluated search engine in the two connection speeds. Google had the highest precision at both SNAL and UDSM Library i.e. 95% and 94% respectively. Yahoo was the second with 94% at SNAL and 88% at

Literature Retrieval Precision of Search Engines



UDSM library. Generally Webcrawler had relatively low precision at both places. The trend of retrieval precision is presented in Figure 1.

MetaCrawler was found to retrieve and index nothing at UDSM library; it was subjected to timeout for every test carried out. This finding was supported by the argument made by Abdeen (1999). This situation may often happen to environments of slow speed like that of UDSM library.

Table 1: *Precision of Search Engines at Sokoine National Agricultural Library (SNAL) and University of Dar es Salaam Library (UDSM).*

Serial No	Search Engine	UDSM Precision %	Mean SNAL Precision %
1	Altavista	69	13
2	Excite	48	16
3	Google	94	95
4	Hotbot	60	53
5	Infoseek	70	67
6	MetaCrawler	00	51
7	Northernlight	55	33
8	SavvySearch	45	59
9	WebCrawler	31	21
10	Yahoo	88	94

Significant tests on precision were done. The null hypothesis (H_0) claimed that there was no difference in terms of precision of search engines at the two connections speeds namely Sokoine National Agricultural Library and University of Dar es Salaam Library. The alternative hypothesis (H_1) in return claimed that there was relatively high precision at Sokoine than at University of Dar es Salaam Library. It was found that at one-tailed test $P(t_0 = -18.97)$ was larger than probability t-distribution $P(t_{0.01, 18} = 2.552)$. This means the results were significantly different at 1% significance level. In other words generally search engines were significantly more precise at the Sokoine National Agricultural Library than at University of Dar es Salaam Library.

3.2 Responsiveness of Search Engines

Responsiveness refers to how quickly the search engines respond from when the query is submitted to the point of displaying the results as well as the time in which the interface loads the hits. Table 2 shows the mean response time at each connection speed. Google had the shortest response mean time at both connections that are 8.09 seconds at University of Dar es Salaam

library and 10.64 seconds at Sokoine National Agricultural Library. MetaCrawler had the longest response mean time at University of Dar es Salaam because the connectivity was relatively slow making it subject to time out.

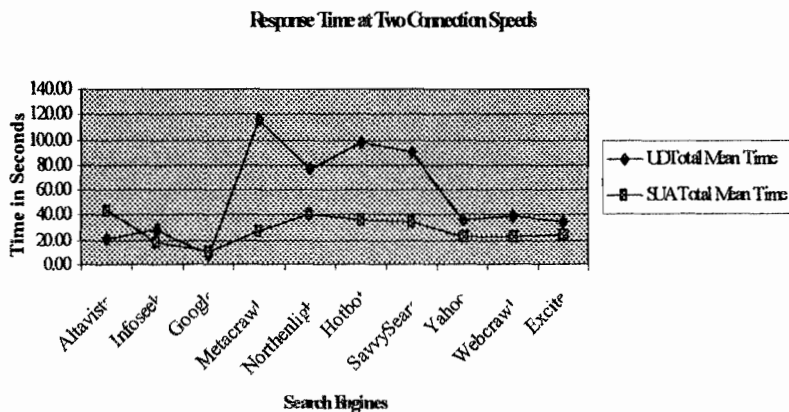


Figure 2

The trend at all connection speeds is graphed in Figure 2 above. It can generally be observed that most of the search engines responded more quickly at Sokoine National Agricultural Library than at University of Dar es Salaam.

Table 2: Response Mean Time of each Search Engine at Sokoine National Agricultural Library and University of Dar es Salaam Library.

Serial No	Search Engine	UDSM Mean Time in SNAL Seconds	Mean Time in Seconds
1	Altavista	21.49	43.61
2	Excite	34.12	24.25
3	Google	08.09	10.64
4	Hotbot	97.19	35.64
5	Infoseek	28.69	18.41
6	Metacrawler	116.20	26.58
7	Northernlight	77.12	41.14
8	SavvySearch	90.08	34.10
9	Webcrawler	39.24	23.21
10	Yahoo	36.42	22.77

After noticing differences in terms of time response at the two connection speeds, making inferences to the whole situation seemed to be essential. The null hypothesis (H_0) assumed that there was no difference in terms of response mean time at Sokoine National Agricultural Library and those at the University of Dar es Salaam. Conversely, the alternative hypothesis (H_1) assumed to have a difference. The test statistics at 5% significance level revealed that generally there was significant difference in terms of time response of search engines at the two connection speeds. A two-tailed test $P(|t_0| = 2.21)$ was larger than probability of t-distribution $P(|t| > 2.10)$. These results are generally taken to be reasonable evidence that the null hypothesis (H_0) was untrue. The implication is that there was a significant difference in terms of responsiveness of search engines at the two connection speeds. In other words search engines responded at significantly higher speeds at the Sokoine National Agricultural Library than at the University of Dar es Salaam Library.

3.3 Evaluation of Other Useful Search Features

Features of ten selected Search Engines were evaluated (see Table 3). The features included the capabilities for phrase search, Boolean search, ability to restrict search to a particular time frame and quality displayed search results. Here are the results and discussion for each search engines.

AltaVista produced poor results on queries that had no Boolean operators i.e. it could bring irrelevant hits that seemed to be highly ranked. Phrase searching for desired results would be improved by use of Boolean search. A query containing a group of words without double quotes or other Boolean operators could index millions of pages most of which were irrelevant. In general AltaVista do not support natural language queries. The ability to restrict search to time frame was moderately supported. The results display contained abstraction with attribution. However, search terms were not obviously seen in bold type making the technical relevance judgement difficult.

Excite supported phrase searching with some difficulties in such a way that it required use of Boolean operators for precise results. In the Advanced search option Boolean searching is in-built. The user is required to select whether it should use "AND", "OR", "NOT" and other operators such as quotation marks, plus (+) and minus (-) signs. Restriction of search to time frame is poorly or not supported. Results are mainly in terms of web sites. Although results may show titles or URL, they are mainly links and web sites. Search terms are not obviously displayed in results.

Google highly supported phrase searching without a requirement of the use of Boolean operators. By using more words in a phrase query it highly improved the relevance of the results. It can be said that using Google in searching you will rarely require many techniques in query formulation. This follows the fact that it ranked highly relevant documents by using normal words without demanding a special syntax. Advanced search option offered an interface, which seemed to have in-built Boolean operators; option could be chosen using radio buttons. The user was required to select the options such as; Find results with ALL of the words, the EXACT PHRASE, with ANY of the words and WITHOUT the words. It showed ability to restrict search to web pages or documents in other formats that had been updated in one year ago. Another feature was to find results updated at any time. Results displays by Google were in an excellent format. Results contained a summary with attribution. The term excellent format in this context implies that search terms (words) appeared in bold simplifying the process for technical relevance evaluation. It was possible to explicitly tell which document was technically relevant. Further more results display contained the format indication. It was easy to identify whether the document was in HTML or PDF formats.

Table: 3 Some features and Searching Mechanism

Search Engine	Advanced Capabilities			Display of Results
	Phrase Search	Boolean Search	Restrict to time frame	
AltaVista	Poor, require Adding "+", "-"	Well Supported	Supported	Summary of Document with Attribution
Excite	Supported	In-built the user	No	Mostly are web sites
Google	Very well supported	In-built flexible well perform	Yes	Excellent summary search terms in bold.
Hotbot	Supported	Well supported	Yes, ten years	Summary with Attribution
Infoseek	Well supported	Supported	No	Summary with Attribution
MetaCrawler	Supported	No	No	Summary with Attribution
Northernlight	Supported	Supported	Yes	Summary with Attribution
SavvySearch	Supported	Poorly supported	No	Summary with Attribution
WebCrawler	Poorly supported	Well supported	No	Web sites or Summary
Yahoo	Well supported	Well supported	Yes, Not more than four years back	Good, summary with attribution search terms appear in bold

Hotbot moderately supported phrase searching in terms of producing highly ranked relevant documents. Boolean search was well supported such that it increased the number of relevant documents and decreased irrelevant

documents. Hotbot showed high capability in restricting search to time frame. It was possible to restrict search to index documents present on the Internet for about ten years ago. The results display was generally not good.

Infoseek well supported phrase queries. However, it brought a limited number of hits that is between 0 to 15 hits. Boolean search was moderately supported because it rarely improved the results. The main weakness with this is that it could not show any ability to restrict search to time frame. The results display contained title summary with attribution. Search terms were not in bold type creating difficulties in direct evaluation for relevancy.

MetaCrawler was expected to bring more desirable results because it is highly recommended in the literature. Not only is it highly recommended but also MetaCrawler is a Meta-Search engine that sends searches to many search engines and display results on single interface.

However in this study, MetaCrawler had a moderate performance, it moderately supported phrase and Boolean search. Restriction to time frame was not supported. Abstraction of result display was found with attribution. Search terms were not bolded hence difficulties in relevance evaluation.

Northernlight supported phrase and Boolean search. It was capable of restricting the search to a specified time frame. Results display consisted of title, summary with attribution. However, search terms were not all shown in bold creating a problem in the process of evaluation relevance.

SavvySearch was among the MetaSearch engine evaluated. It also revealed moderate performance. It supported phrase search. Boolean search was poorly supported. It did not show ability to restrict search to time frame. Search results were displayed with the title link, abstract of the document found with attribution. A common problem was that the search terms were not obviously indicated in bold type.

WebCrawler supported phrase searching. However, high numbers of irrelevant materials were indexed. Boolean search could reduce the number of irrelevant materials. Restriction of search to a specific time frame was not evidently noticed. The hits produced were mainly leading a web site not the particular document as in other search engine. This situation was also evident with Excite.

Yahoo well supported both phrase and Boolean search. It manifested ability to restrict search to a specific time frame - not more than four years ago. Results display was generally good consisting of title, abstract with

attribution. An interesting quality of result display was that search terms were in bold reducing the difficulties of evaluating the relevance of indexed materials.

Summary and Recommendations

4.1 Summary

In terms of relevance and precision Google outperformed all search engines scoring the highest precision at both SNAL and UDSM Library i.e. 95% and 94% respectively. Yahoo was the second with 94% at SNAL and 88% at UDSM library. The results were significantly different at 1% significance level. In other words generally search engines were significantly more precise at the Sokoine National Agricultural Library than at University of Dar es Salaam Library. There is an indication that search engines are more precise on higher connection speeds. Meta Crawler was found to retrieve nothing at UDSM library; it was subject to timeout for every test carried out.

Google had the shortest response mean time at both connections that are 8.09 seconds at University of Dar es Salaam library and 10.64 seconds at Sokoine National Agricultural Library. The test statistics at 5% significance level revealed that generally there was significant difference in terms of time response of search engines at the two connection speeds. MetaCrawler had the longest response mean time at University of Dar es Salaam because the network performance speed was relatively slow subjecting it to time out. This means, metasearch engines like metacrawler seemed to perform poorly under slow connection speeds.

Google supported exceedingly well the phrase searching than others. Yahoo was the second in this aspect. MetaCrawler and Savvy Search did not or supported poorly Boolean searching. Most of the Search engines had no ability to restrict search to a specific time frame. The result displays were excellent for Google followed by Yahoo.

4.2 Recommendations

Google and Yahoo are therefore recommended to be given higher priority in searching for literature until they prove otherwise. However, it should be noted that it is not a guarantee to find what you need. Therefore try as many search engines as you can to meet your information need. This recommendation is true for Agricultural subjects which were used in the experiments.

More studies are recommended especially to evaluate search engines under slow connection speeds. Relatively more intensive studies with more search engines, more search queries from different subject areas and more connection speeds are highly recommended.

There is an indication that some search engines perform poorly in slow speed contrary to the literature recommendation that they are able to cover

wide search area (Repman and Carlson 1999). For example MetaCrawler is said to be a metasearch engine that uses other search engines to search for information hence more search coverage. However, it was not able to retrieve any document at the University for Dar es Salaam Library. In this situation users are advised not to depend on them because they might not provide the desired results.

Finally, it is also recommended that the connection speed be improved to allow for smooth use of search engines in our literature search.

References

- Abideen, P.S. (1999) *Meta Search Engines: effective tool for information retrieval*. CALIBER 99. Academic Libraries in the Internet era.: Proceedings of the sixth National Convention for automation of libraries in education and research. Nagpur, India 18-20 Feb, 1999
- Allen, B. L (2001) *Boolean browsing in an information system. An experimental test*. Information Technology and libraries. Vol. 20. No.1. pp12-20
- Bar-Ilan, J (2000) *Evaluating the stability of the search tools Hotpot and Snap: A case study*. Online Information Review Vol. 24 No. 6
- Becket, D (1998) *Search Engine Corner*.
<http://www.ariadne.ac.uk/issue16/search>
- Chatfield, C (1983) *Statistics for technology*. A cause in applied statistics. 3rd ed. Chapman and Hall, New York.
- Chu, H and Rosenthal, M (1996) *Search engines for the World Wide Web : a comparative study and evaluation methodology ASIS96*,
http://www.asis.org/annual_96/electronic_proceedings/chu.html
- De Smet, E and Kerstens, V (2000) *Supply of Academic publications (SAP): Towards a sustainable electronic document delivery system for southern libraries using the Internet*. A report for SAP project.
- Dong, X and Su, L.T (1997) *Search engines on the World Wide Web and information retrieval from the Internet: a review and evaluation*. Online and CD-ROM review. Vol. 21 no.12 pp.67-81.
- Hanka, R and Fuka, K (2000) *Information overload and "Just in time" knowledge*. Electronic Library Vol. 18 No. 4 pp 279-284.
- Introna, L and Nissenbaum, H (2000) *Shaping the Web: why the politics of search engines matters*. Information Society Vol.16 no.3 pp.167-185
- Katundu, D. R (1998) *The use and sustainability of Information Technology in academic and research libraries in Tanzania*. A thesis submitted as a fulfilment of PhD.
- Kibirige, H. M and De Palo, L. (2000) *The Internet as a source of Academic Research Information: Finding of two pilot studies*. Information Technology and Libraries. Vol. 19 no. 1 p. 11-16.
- Lawrence, S and Giles, C. L (1998) *Searching the World Wide Web*. Science vol. 280 pp. 98-100.
- Lawrence, S (2001) *Online or invisible?* Nature vol.411, no. 6837, pp. 521.
- Lesser, V. et al (2000) *BIG: an agent for resource-bounded information gathering and decision-making*. Artificial Intelligence Vol. 118 no. 1-2 pp197-244.
- Levey, L.A (2000) *Wired for Information: putting the Internet to good use in Africa*. A Project for Information Access and connectivity(PIAC), Nairobi.

- Mann, P. S. (1995) *introductory statistics*. 2nd Edition. John Wiley and sons, Inc., New York.
- MetaCrawler (2001) *MetaCrawler fact sheet*. <http://www.metacrawler.com/>
- Mnyani, Z.V.G (2000) *Factors influencing effective application and use of information and communication technology in Tanzania*. A dissertation submitted at the University of Dar es salaam as a partial fulfilment of masters of arts in information studies.
- Niewenhuysen, P. (1999) *Subjects for research or study*. Vrije universities Brussels. Brussels.
- Proctor, L (1997) *Academic Research on the Internet*. <http://web.uvic.ca/comped/online/research/>
- Repman, J and Carlson, R. D (1999) *surviving the storm: Using Metasearch Engines Effectively*. Computers in Libraries. Vol. 19 no. 5.
- Salton, G (1989) *Automatic Text processing*. Addison-Wesley, Reading, M.A
- Sasikala, C and Patnaik, K.R (1999) *A comparative study of two web search engines: Altavista and Excite*. Proceedings of the sixth national convention for automation of libraries in education and research, Nagpur. Pp 346-54.
- Tomaiuolo, N.G. and Packer, J. G. (1996) *Web search engines: key to location information for all users or only the cognoscenti?* Online information96. Proceedings of twentieth International online information meeting. London 3-5 Dec. 19996. P 41-8.
- Tweve, J. T (2000) *An investigation of the availability and application of Information technology in the institute of higher learning*. A dissertation submitted at the University of Dar es salaam as partial fulfilment of Masters of Arts in information studies.
- Vidmar, D.J. (1999) *Darwin of the web: The evolution of search tools*. Computers in Libraries, vol.19, No.5, pp 23-29.
- Weimer, M.L and Rusch, P.F (1996) *New searching technologies and interface online information 96*. Proceedings of the twentieth International online information meeting. pp221- 4.
- Wema, E.F (2000) *The impact of introducing computers into library services*. A dissertation submitted at the University of Dar es salaam as partial fulfilment of masters of Arts in information studies.