Precision, Recall and F-Measure of Select Search Engines: Retrieval of Research Article in the Field of Library and Information Science

Peerzada Mohammad Iqbal*, Abdul Majid Baba, Bilal Ahmad Bhat*****

Abstract

This paper presents the results of a research conducted on seven search engines - Google, Bing, Yahoo, Ask, Baidu, Dogpile and DuckDuckgo for retrieving scholarly articles using Library and Information Science related search terms. The search engines are evaluated by taking the first twenty results pertaining to 'scholarly articles' for estimation of precision, recall and F-measure. It shows that Baidu is most comprehensive in retrieving 'scholarly articles' followed by Google and Dogpile. It also reveals that all search engines (except Ask and Dogpile) perform well on structured queries while Ask and Dogpile performs better on unstructured queries.

Keywords: Search engine, Precision, Recall, F-measure, Scholarly information, Library and Information Science

Introduction

The World Wide Web can be used as a quick and direct reference to get any type of information in electronic format all over the world. However, information found on the Web needs to be filtered and may include voluminous misinformation or non relevant information. The user or Internet surfer may not be aware of quality search engines to get information on a topic quickly and may use different search strategies. Finding useful information quickly on the Internet poses a challenge to both the ordinary users and the information professionals. Though the performance of currently available search engines has been improving continuously with powerful search capabilities of various types, the lack of comprehensive coverage, the inability to predict the quality of retrieved results, and the absence of controlled vocabularies make it difficult for users to use search engines effectively. The use of the Internet as an information resource needs to be carefully evaluated as no traditional quality standards or control have been applied to the Web. Librarians need to be able to provide informative recommendations to their clientele regarding the selection of search engines and their effective search strategies.

Related Literature

Internet has become a part of our daily life as the Web is now a necessary tool for collecting information and other online activities. Search engines provide convenience in information retrieval because they collect information from many different web sites and provide a gist of information with web addresses. Search engines have emerged during the last two decade of the 20th century. Though the nature of search engines is quite recent but the body work on web search engines is already quite sizable and is expanding exponentially. Nowadays, most of the literature is available in electronic format which describes various attempts at conducting evaluation of web search engines. Several parameters like precision, recall, accessibility, coverage etc have been used in reporting the findings. The following paragraphs take an overview of such studies.

Cleverdon and Keen (1966) listed six criteria that could be used to evaluate an information retrieval system:

(1) Coverage	(2) Time Lag
(3) Recall	(4) Precision
(5) Presentation and	(6) User Effort

Of these criteria, recall and precision have most frequently been applied in measuring information retrieval. Both of these measures have been criticized

* Senior Professional Assistant, Sher-e-Kashmir University of Agricultural, Sciences and Technology of Kashmir (SKUAST-K), Jammu & Kashmir, India. Email: iqlis2007@gmail.com

** Head, Department of Library & Information Science, University of Kashmir, Srinagar, Jammu & Kashmir, India. Email: abmbaba@kashmiruniversity.ac.in; abmbaba1@rediffmail.com

*** Sr. Assistant Professor, Department of Statistics, Faculty of Fisheries, Sher-e-Kashmir University of Agricultural, Sciences and Technology of Kashmir (SKUAST-K), Jammu & Kashmir, India. Email: bhat_bilal@rediffmail.com

for a variety of reasons, and range alternatives have been suggested. However, in spite of their deficiencies, recall and precision are continued to be used widely, in part because of insufficient consensus about which alternative measures might be superior. Chu and Rosenthal (2003) evaluated the capabilities of Alta Vista, Excite, and Lycos in terms of retrieval performance measured by response time and precision calculated for the first ten hits. Their study also proposed a methodology for evaluating WWW search engines in terms of five aspects:

Clarke and Willett (1997) reported a comparison of the retrieval effectiveness of AltaVista, Excite and Lycos using recall, precision and coverage for thirty queries searched using the three search engines. Oppenheim, Morris, McKnight, and Lowley (2000) criticize on the evaluation of web search engines with the aim of identifying some of the issues and problems in this rapidly dynamic field. On the other hand Monica and Steven (2000) have given a critical and in depth analysis of information retrieval techniques for evaluation of search engines to highlight the importance of a scientifically sounded approach to search engine evaluation. The author modify the use of classic evaluation techniques "relevance" "recall" and "precision" a set of flexible and proceeding guidelines for the use of these techniques are identified. Courtoius and Berry (1999) tested five major search engines (AltaVista, Hotbot, Excite, Infoseek and Lycos) for understanding retrieval and ranking of documents for sample search queries. Chu and Rosenthal (2003) evaluated and compared three web search engines (AltaVista, Excite and Lycos) to check their search capabilities such as use of Boolean logic, truncation, field search, word or phrase search and for measuring precision. The authors used ten search queries drawn from real reference questions and using available command features for each engine. Bar-Ilan (1998) argues by investing retrieval capabilities of six search engines on a single query. As a case study the query "Erosos" was chosen. The author also reported that search engines do not perform well in complex queries involving more than one search terms.

Mettrop and Nieuwenhuysen (2001) have made a practical investigation of thirteen search engines including AltaVista, Excite, Hotbot, Lycos and Infoseek for gauging the degree of consistency to which an engine retrieves documents. Shafi and Rather (2005) conducted a research on five search engines (AltaVista, Google, HotBot, Scirus and Bioweb) using Biotechnology related search terms for retrieving scholarly publications. The first ten results on every search engine were evaluated pertaining to 'scholarly information' for estimation of precision and recall. Their study shows that Scirus is most comprehensive in retrieving 'scholarly information' followed by Google then HotBot. While as Kumar and Prakash (2009) compared the retrieval effectiveness of the Google and Yahoo, both precision and relative recall was considered for evaluation. Queries were divided into one-word, simple multi-word and complex multi-word using concepts in the field of library and information. Besides, a few investigators have attempted to evaluate search engines for measuring precision, recall and F - measure. The attempt of Sudhakar, Poonkuzhali and Kumar (2012) and Mariappan, Suresh & Bharathi (2013) are worth mentioning.

Sudhakar, Poonkuzhali and Kumar (2012) have proposed content based ranking for search engines by using parameters of precision, recall and F-Measures to text based searching. In support to this proposal Mariappan, Suresh & Bharathi (2013) also suggests semantic measure for obtaining the relevant query to search through different search engines. Their experimentation part was performed with different set of queries and the performance of the results was analyzed with precision, recall and F-measure. From the results, it was found that the proposed meta search engine has performed better than existing work by achieving the precision of 0.8.

Although many studies are carried out there has been little consistency in the way such studies were carried out. The reported findings obviously do not appear to agree with one another. The methodologies and evaluation criteria used by these studies differed as well. This review demonstrates that only a few studies have tried to evaluate the performance of search engines using systematic measures of effectiveness. Most studies have employed precision and have made little or no attempt to assess the recall performance of various search engines and no sign of F-measure.

Problem

In the beginning of the internet, it was easy to fine information using variety of software that was usually command driven rather than using a graphical interface. With the proliferation of information, systems such as Archie, Gopher and Veronica became increasingly unable to cope with huge information. The advent of many types of search engines provided solution for literature search using Boolean operators, Proximity searching, Wild cards, Truncation etc. Many search engines developed new versions and techniques to achieve some kind of sophistication but all have not helped to forward the case of

access and searching from scholar's perspective. Besides keeping in view different ways of indexing the internet, search engines operate in different ways and retrieve documents in different orders. Further, it does not sift information form scholar's point of view i.e., it retrieves information on a particular topic from different aspects like marketing, advertisement, news and entertainment mixed with some research papers. The academic community attempts to look purely for scholarly information on his topic of interest to have output / retrieval best in terms of recall, precision, comprehensiveness and devoid of fluctuations etc.

The present investigation attempts to evaluate the performance of the select search engines in retrieving scholarly research articles in the field of Library and Information science with respect to their richness in databases, currency, repetition, fluctuations, dead Links, recall, precision and F-measure.

Rationalize/Hypothesis

For the study two hypotheses are formulated;

H1: (Null Hypothesis)

"The recall, precision as well as F-measure of search engines does not differ significantly"

H2: (Alternative Hypothesis)

"The recall, precision as well as F-measure of search engines does differ significantly"

The above hypotheses are supported by Shafi & Rather (2005) used Friedman's two way analysis of variance. Precision is the probability that a (randomly selected) retrieved document is relevant. Recall is the probability that a (randomly selected) relevant document is retrieved in a search.

The study is interested in determining whether the selected search engines are significantly different in their performances precision, recall and F-Measure test set. To conclude that these selected search engines are significantly different, a null hypothesis and an alternative hypothesis is formulated for the study that can be determined by the work according to the database they create on entering specific search terms from end user.

The selected search engines which are to be evaluated may overlap their results; while some of them will be included for the pooling. The process can be illustrated by taking these search engines (Say a, b, c, d, … n) into consideration which may retrieve (a1, b1, c1, d1, ….. n1) scholarly articles respectively. Further assuming that there is no overlap between search engines

i.e., $a \cap b \cap c \cap d \cap \ldots\ldots n = \text{zero}$,

Then the relative recall of search engine 'a' is

Again if overlapping exists between search engines

i.e., $a \cap b \cap c \cap d \cap \ldots.. n = a2$

Then the relative recall of engine 'a' is

Objectives

The following objectives are laid down for the study:

- To select search engines.
- To select search terms for the study.
- To assess the precision, recall and F - measure of "scholarly documents" retrievable via select search engines.

Method

As certified by International Standard Organization there are 230 search engines (Promote3.com, 2015) available for searching the web. These search engines are of various types like general search engine, robotic search engine, Meta search engine, directories and specialized search engines. Most users prefer robotic search engines as they allow the users to compose their own quires rather than simply follow pre specified search paths or hierarchy as in case of directories. Moreover, robotic search engines locate data in a similar way i.e., by the use of crawlers or worms. This distinguishing feature differentiates them form web directories like Yahoo! Where collections of links to retrieve URL's are created and maintained by subject experts or by means of some automated indexing process. However some of these services are also include a robot driven search engine facility. But this is not their primary purposes. This due to this feature Yahoo! Was included for the study.

Meta search engine e.g., Dogpile etc don't have their own database. These access the database of many robotic search engines simultaneously. Thus these were included for the study.

Still hundreds of robotic general search engines navigate the web, in order to limit the scope of study after preliminary study, following criteria was laid down for selection of general search engines:-

a) Availability of automated indexing
b) Global coverage to data
c) Quick response time
d) Availability of filter search mechanism
e) Least overlapping
f) Major market holder

Following two general search engines were selected for the study for meeting all the criteria and being comprehensive in nature.

a) Google b) Baidu

Since the study relates to the field of Library and Information Science. It was felt to include specialized search engine in the study representing question answer search engine i.e., Ask.com & another specific i.e., Bing.

There being no full-fledged search engine in the field Library and Information Science except many associated with library websites. Among those human powered (DuckDuckGo) after preliminary investigation and feasibility in the study was included in the study. Thus the search engines undertaken for evaluation of study are:-

a) Google (General)
b) Bing (Specific)
c) Yahoo! (Directory)
d) Ask (Question Answer Search engine)
e) Baidu (Country Specific General Search engine)
f) Dogpile (Meta search Engine)
g) DuckDuckgo (Human Powered Search Engine).

Selection of Terms

Selection of terms is not directly possible in development and multidimensional field like Library and Information Science. Therefore, classification schemes like DDC (18th) and DDC (22nd) were consulted to understand Broad/ Narrow structure of Library and Information Science. It helped to get five terms/Fields i.e.,

a) Information System
b) Digital Library
c) Library Automation
d) Library Services
e) Librarianship

These terms were then browsed in "LC list of subject Headings" which provided many other related terms (RT) and Narrow terms (NT). Further NT and RT attached to each other preferred or standard terms were also browsed which retrieve a large number of Library and Information Science terms. At first instance 140 Library and Information Science related terms were identified.

Some terms occurred more than once and duplication removed. It reduced the number to 100. Later terms were divided into three broad groups under:

a) Application
b) Transformation
c) Inter-relation

"Application" denotes utility of Library and Information science in various fields and about 50 terms came under this group. "Transformation" refers to a method of developing or manufacturing library services into practical market and 30 terms fall under this group. "Inter-relation" means transformation/dependence of one subject onto another and 20 terms came under this group.

Further each category is sub-divided into groups.

"Application" into four i.e., "Reference service", "Informatics", "Information Retrieval" & "Information Sources". "Transformation" into two i.e., "Digitization" & "Consortia". "Inter-relation" into two i.e., "Library Network" & "Information System".

The terms in each group were arranged alphabetically and each term was given a tag. Later 19% of the terms were selected from each group using "Systematic Sampling" (i.e., first item selected randomly and next item after specific intervals). It further reduced the number to 19. Finally the selected terms were classified into three groups under "Simple", "Compound" & "Complex Terms" (Fig:- 1.1). This was done in order to investigate how search engines control and handle simple and phrased terms.

"Simple Terms" containing a single word were submitted to the search engine in the natural form i.e., without punctuating marks. "Compound Terms" consisting of two words were submitted to the search engines in the form of phrases as suggested by respective search engines and "Complex Terms" composed of more than two words or phrases, were sent to the search engine with suitable

Boolean operator "AND" & "OR" between the terms to perform special searches.

S. No	*Simple terms*	*Compound Terms*	*Complex Terms*
1	Catchwork	Bibliometric Classification	Digital Library Open Source Software
2	Citation	Citation Analysis	Health Information System
3	Dublincore	Comparative Librarianship	Library Information System
4	Indexing	Digital Preservation	Library Information Network
5	Manuscript	Electronic Repositories	Multimedia Information Retrieval
6	Plagiarism	Library Automation	
7	Reprints	Semantic web	

Fig: 1: Keywords

Selection of Search Results and Filteration Technique

To evaluate the select search engines top 20 results from each search engine was taken into consideration to determine precision. The assessment of top 20 results is supported by Hawking, Craswell, Bailey & Griffiths, (2001) compared 20 search engines using first top 20 search results comparing 54 topics originated by anonymous searchers for measuring search engine qualities. Tongchim, Sornlertlamvanich & Isahara, (2006) used seven search engines for measuring effectiveness of search engines on Thai queries. Their results calculated from binary relevance judgments of the first 20 returned results, using 56 topics. Latter Egelman, S., Cranor, L., & Chowdhury, A. (2006) conducted a study of quality and quantity of (Platform for privacy preferred project) P3P-encoded policies associated with top-20 search results from three popular search engines viz., AOL, Google, and Yahoo!. The study examined top 20 search results returned by each search engine to build a P3P-enabled search engine and used it to gather statistics on P3P adoption as well as the privacy landscape of the Internet as a whole. Dirk **(2008)** Evaluated 5 search engines using first top 20 hits for retrieval effectiveness of web search engines.

Andago, Phoebe & Thanoun, (2010) collected queries from 30 university students and entered these queries into two search engines viz., Google and Hakia. Precision was thereafter calculated using first 20 hits. The 20 results were taken into evaluated for a comparison of precision of Semantic Search Engine against a Keyword Search Engine. Ajayi & Elegbeleye (2014) used first 20 results for performance evaluation of three search engines. The use of first 20 results were thought to be genuine as majority of scholars use first two pages of search hits which by default to many search engines is fixed to 10 hits per page.

The evaluation of first top 20 hits was further backed by a questioner among the scholars of Kashmir University. A total of 100 questioner were distributed among the doctorial and M.Phil scholar of said university. The aim was to check the result extension at maximum and type of filtration a scholar uses. It was revealed that 84 percent of the scholar prefer first 20 hits (or two pages: a default of 10 results per page), 11 percent prefer first 10 results and five percent prefer more than 20 results. Further it was revealed that scholar's use PDF (portable document format) to filter the results as to get maximum of research article.

Pointing Scale

To determine the relevance of each page, a five point Likert scale was used which enabled the calculation of precision. Likert is a psychometric scale primarily used in questionnaires to obtain preferences or degree of agreement with a statement or set of statements. Likert scale is unidimensional (Measures a single trait) and non comparative in nature. The scale describes level of agreement with given or profound statements by way of an ordinal scale (Bertram, 2016). Likert scale was conceived back in 1929 by Gardner Murphy (Likert, 1932). The writer firstly aimed at five major problems to present a wide array data as in attitude area:-

a) International relations,
b) Race relations,
c) Economic conflict,
d) Political conflict, and
e) Religion.

Four classes of questionnaire material fall into this category:

a) Yes-no,

b) Multiple choice,

c) Propositions to be responded by degrees of approval, and

d) A series of brief newspaper narratives to be approved or disapproved in various degrees.

Likert scale is widely used by Extension professionals. Many researches and scholars are using Likert scale for research analysis in a random sampling environment. Shafi & Rather, (2005) used 4 point Likert technique for retrieving precision and recall of 5 major search engines in the field of Biotechnology. In 2010, Journal of Extension published at least 21 articles where Likert technique was used.

In 2011 same journal accepted publications using 4-point Likert alternatives (Behnke & Kelly, 2011; Robinson & Shepard, 2011), five-point Likert alternatives (Diker, Walters, Cunningham-Sabo, & Baker, 2011; Elizer, 2011; Hines, Hansen, & Falen, 2011; Kalambokidia, 2011; Kroth & Peutz, 2011; Singletary, Emm, & Hill, 2011), six-point Likert alternatives (Allen, Varner, & Sallee, 2011; Beaudreault & Miller, 2011; Wyman et al., 2011), and a seven-point Likert alternative (Walker, Vaught, Walker, & Nusz, 2011).

Keeping in view the widely accepted format of pointing scale, a five point scale was employed for the study as under:

- The page representing full text of research paper, seminar/conference proceedings or a patent is given a score 4.
- The page corresponding to an abstract of a research paper, seminar/conference proceedings, slides of research dissertation, tutorials or a patent is given a score of 3.
- The page corresponding to a book, database or library related licensed software is given a score of 2.
- The page representing other than the above (i.e., Wikipedia, Company web pages, Dictionaries, Encyclopedia, Organization, Bulletin Boards, Social Sites related to concerned subjects, general related videos etc) is given a score of 1.
- If a page occurs more than once under different URL's is assigned a score of Zero.
- If a page is irrelevant is assigned a score of Zero (an advertising page, a page of no value in relation to the given keywords).
- In a situation where server didn't respond or "File not Found" occurred is assigned a score of Zero (after three attempts).

Results and Discussion

Relative Recall & Precision

Table 1: Kruskal-Wallis Test for Precision

Name of the Search Engine	*No. of Observations*	*Mean Rank*	***H*(X^2) *test* *(Chi Square)***	*Significance*
Google	19	72.97	5.048	.538
Bing	19	62.66		
Yahoo	19	63.97		
Ask	19	72.74		
Baidu	19	77.95		
Dogpile	19	64.87		
Duck-DuckGo	19	53.84		

Significance < 0.05, Degree of freedom (df) = 6, Median=1.5000

The data presented in Table 1 shows the mean rank of select search engines for overall precision. It is observed that highest mean rank is shown by Baidu (77.95), followed by Google (72.97). Ask also shows a good mean rank of (72.74) followed by Dogpile (64.87). Yahoo! and Bing show a mean rank of (63.97) and (62.66) respectively. DuckDuckGo shows the least mean rank of (53.84). Statistically the mean rank for precision does not differ significantly at 0.05 level of significance. Thus the interference can be drawn that precision of search engines also does not differ significantly. Hence null hypothesis H1 is accepted. However the alternative hypothesis (H2) is rejected at 0.05 level of significance.

Table 2: Kruskal-Wallis Test for Relative Recall

Name of the Search Engine	*No. of Observations*	*Mean Rank*	$H(X^2)^{test}$ ***(Chi Square)***	*Significance*
Google	19	55.92	24.870	.000
Bing	19	56.24		
Yahoo	19	63.03		
Ask	19	71.76		
Baidu	19	101.63		
Dogpile	19	73.39		
DuckDuckGo	19	47.03		

Significance < 0.05, Degree of freedom (df) = 6, Median=0.1300

From Table 2 it is evident that the mean rank of Baidu (101.63) is highest for relative recall among select search engines. It is observed that 2nd highest mean rank for relative recall is shown by Dogpile (73.39). Ask shows a mean rank of (71.76) while Bing and Google shows mean rank of (56.24) and (55.92). DuckDuckGo shows the least mean rank among the search engines (47.03). Statistically there is a significant difference among select search engine in terms of Recall (i.e., $p<0.05$). Thus the interference can be drawn that Recall of search engines differ significantly. Hence H2 is accepted. However the null hypothesis (H1) is rejected at 0.05 level of significance.

a) Baidu has the highest relative recall (0.34) for scholarly documents in the field of Library and Information Science followed by Dogpile (0.15). Yahoo! and Ask ranks third (0.12) respectively. While Google (0.10) and Bing (0.10) the fourth position whereas DuckDuckGo has the least recall value (0.08).

b) Among general search engines Baidu surpass Google in relative recall even though Google retrieves greater number of comprehensiveness. This is due to Baidu's own scholar search (known as Xeshu.Baidu.com) which retrieved more scholarly hits than Google.

c) Baidu has highest precision for scholarly documents (1.79) in the field of Library and Information Science followed by Ask (1.75). Google ranks third (1.73) followed by Dogpile (1.68) whereas Bing and Yahoo! (1.56) scores at equal. DuckDuckGo scores the least precision (1.42).

d) Baidu ranks first in both the respects (i.e., relative recall and precision). Ask score second rank in achieving precision and forth in relative recall whereas Google ranks third in precision and fifth in relative recall. Dogpile scores forth in precision and second in relative recall. Bing attains fifth position in precision and sixth in relative recall while as Yahoo! scores sixth in precision and third in relative recall where as DuckDuckGo retains seventh position both in relative recall and precision.

The precision is inversely proportional to relative recall i.e., if precision increases recall decreases and vice versa.

Use of complex and compound queries enhance precision and relative recall of search engines i.e., more specific a query more relevance or precision it will achieve.

F-measure

Search engines	*F-score*
Baidu	0.57
Dogpile	0.28
Ask	0.23
Yahoo!	0.22
Bing	0.19
Google	0.18
DuckDuckGo	0.15

Fig. 2: F-Score of all Search Engines

F-measure is a weighed mean for both precision and recall giving a balanced measure for harmonic mean from both the values. Table 2 shows Baidu as the highest scorer value (0.57) followed by Dogpile (0.28) attaining a second position while as Ask scores a third rank (0.23). Yahoo! attains forth rank with F-score of (0.22) whereas Bing scores (0.19) scoring a fifth rank. Google shockingly

ranks at sixth with a score of 0.18 while as DuckDuckGo attains a rank of seventh with a score of 0.15.

Searching Capacity

These findings are based on observations, experience and literature survey.

a) The indexing mechanism of search engine varies from engine to engine which changes for each search engine over time. The new techniques lead to loss of documents that one wishes to retrieve at different intervals of time.

b) Ask, Dogpile and DuckDuckGo lacks comprehensiveness in their databases and doesn't have a hit meter. Dogpile being a Meta search engine lacks in consistency in its indexing pattern. DuckDuckGo is a human powered search engine shows a good indexing mechanism but also lacks comprehensiveness in its database same as Ask search engine.

c) Baidu is a country specific search engine and in Chinese language but supports foreign language options too (English). Baidu has versatility in retrieving scholarly documents in the field of library and information science.

d) Search Engine giant Google being a large comprehensive search engine shows repetition in first pages of results. Often it was learnt that the engine shows multiple similar results by different URL's.

Conclusion

The immense size and constant transformation of web has made it impossible for search engines to retrieve relevant information. To overcome the problem a variety of search tools ranging from general to subject specific search engines have been developed to provide considerable assistance for finding relevant information on the web. Various studies have been carried out to evaluate search engines for their performance and identify problems associated with them. But due to dynamic nature of search engines each study will not prove final over time.

This study (different form previous studies) attempted to find the best search engine suitable for locating scholarly documents in the field of library and information science. The study has no inclination towards any search engine as results are purely based on evaluation of the results. The study reveals that those who intend to find current scholarly documents in the field of library and information science, Baidu provides the best opportunity especially for those who have direct access to various online journals or consortiums etc. Google is the best alternative for getting web resources and its "Google Scholar" for accessing scholarly information has borne new dividends for the researches. Baidu shows the highest recall and precision due to the induction of its journal citations along with web sources, otherwise Google would rank the highest. Ask offer good combination of currency and comprehensiveness along with precision and relative recall. Thus Bing deserves to be recommended along with Baidu and Google. Yahoo once prominent on the web has lagged behind. DuckDuckGo is the poorest among the search engines in all respects. The results do reveal that the select search engines associate fluctuations which need to be taken into consideration while searching the web.

References

Ajayi, O. O., & Elegbeleye, D. M. (2014). Performance Evaluation of selected search engines. *Computer Engineering and Intelligent Systems*, 5(1), Retrieved from http://www.iiste.org/Journals/index.php/CEIS/article/viewFile/10301/10504

Allen, K., Varner, K., & Sallee, J. (2011). Addressing nature deficit disorder through primitive camping experiences. *Journal of Extension*, 49(3). Retrieved from http://www.joe.org/joe/2011june/iw2.php

Andago, M. O., Phoebe, T., & Thanoun, B. A. M. (2010). Evaluation of a semantic search engine against a keyword search engine using first 20 precision. *International Journal for the Advancement of Science & Arts*, 1(2), 55-63. Retrieved from http://www.ucsiuniversity.edu.my/cervie/pdf/ijasa/paper-V1N2IT3.pdf

Bar-Ilan, J. (1998). On the overlap, the precision and estimated recall of search engines: A case study of the query "Erdos". *Scientometrics*, 42(2), 207-208. DOI: 10.1007/BF02458356

Beaudreault, A. R., & Miller, L. E. (2011). Need for methamphetamine programming in Extension education. *Journal of Extension*, 49(3). Retrieved from http://www.joe.org/joe/2011june/rb6.php

Behnke, A. O., & Kelly, C. (2011). Creating programs to help Latino youth thrive at school: The influence of Latino parent involvement programs. *Journal of Extension*, 49(1). Retrieved from http://www.joe.org/joe/2011february/a7.php

Bertram, D. (2016). Likert Scales. The faculty of mathematics: *University of Belgrade*, Article CPSC

-681. Retrieved from http://poincare.matf.bg.ac.rs/~kristina/topic-dane-likert.pdf

Chu, H., & Rosenthal, M. (2003). Search engines for the World Wide Web: a comparative study and evaluation methodology. *Information* Today, 127-135. Retrieved from http://cui.unige.ch/tcs/cours/algoweb/2002/articles/art_habashi_arash.pdf

Clarke, S. J., & Willett. P. (1997). Estimating The Recall Performance of Web Search Engines. *Aslib Proceedings*, 49(7), 184-189. DOI: 10.1108/eb051463

Cleverdon, C. W., Mills, J., & Keen, E. M. (1966). An inquiry in testing of information retrieval systems, 2. Cranfileld, U.K.: Aslib Cranfield Research Project, College of Aeronautics. Retrieved from http://www.imedia.mie.utoronto.ca/files/webIR_eval1_99.pdf

Courtois, M. P., & Berry, M. W. (1999). Results Ranking in Web Search Engines. *Online*, 23(3), 39-46.

Diker, A., Walters, L. M., Cunningham-Sabo, L., & Baker, S. S. (2011). Factors influencing adoption and implementation of cooking with kids: An experiential school-based nutrition education curriculum. *Journal of Extension*, 49(1). Retrieved from http://www.joe.org/joe/2011february/a6.php

Dirk, L. (2008). The retrieval effectiveness of web search engines: Considering results descriptions. *Journal of Documentation*, 64(6), 915-937. DOI: 10.1108/00220410810912451

Egelman, S., Cranor, L., & Chowdhury, A. (2006). *An analysis of p3p-enabled web sites among top-20 search results*. In proceedings of the 8th International Conference on Electronic Commerce (pp. 197-207). Retrieved from http://casos.cs.cmu.edu/publications/papers/icec06.pdf

Elizer, A. H. (2011). Are transformational directors required for satisfied agents? *Journal of Extension*, 49(2). Retrieved from http://www.joe.org/joe/2011april/rb1.php

Hawking, D., Craswell, N., Bailey, P., & Griffiths, K. (2001). Measuring Search Engine Quality. *Information Retrieval*, 4(1), 33-59. DOI: 10.1023/A:1011468107287

Hines, S. L., Hansen, L., & Falen, C. (2011). So, you want to move out? An awareness program of the real costs of moving away from home. *Journal of Extension*, 49(1). Retrieved from http://www.joe.org/joe/2011february/iw2.php

Kalambokidia, L. (2011). Spreading the word about extension's public value. *Journal of Extension*, 49(2). Retrieved from http://www.joe.org/joe/2011april/a1.php

Kroth, M., & Peutz, J. (2011). Workplace issues in extension - A Delphi study of extension educators. *Journal of Extension*, 49(1). Retrieved from http://www.joe.org/joe/2011february/rb1.php

Kumar, B. T. S., & Prakash, J. N. (2009). Precision and relative recall of search engines: A comparative study of Google and yahoo. *Singapore Journal of Library & Information Management*, 38. Retrieved from http://www.las.org.sg/sjlim/SJLIM20094Sampath.pdf

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22(140), 1-55. Retrieved from http://psycnet.apa.org/psycinfo/1933-01885-001

Mariappan, A. K., Suresh, R. M., & Bharathi, V. S. (2013). Semantic meta search engine using semantic similarity measure. *Journal of Theoretical and Applied Information Technology*. 50(3). Retrieved from http://www.jatit.org/volumes/Vol50No3/2Vol50No3.pdf

Mettrop, W., & Nieuwenhuysen, P. (2001). Internet search engines-Fluctuations in document accessibility. *Journal of Documentation*, 57(5), 623-651. DOI: 10.1108/EUM0000000007096

Monica, L., & Steven, B. (2000). Information Retrieval Techniques for Evaluating Search Engines: A Critical Overview. *Aslib Proceedings*, 52(3), 124-129. DOI: 10.1108/EUM0000000007006

Oppenheim C., Morris A., McKnight C., & Lowley S. (2000). The Evaluation of WWW Search Engines. Journal of Documentation, 56(2), 190-211. DOI: 10.1108/00220410010803810

Promote3.com (2015). *Top Search Engine Ranking Search Engine Optimization*. IDV International: California. Retrieved from http://www.promote3.com/search-engine-230.htm

Robinson, P., & Shepard, R. (2011). Outreach, applied research, and management needs for Wisconsin's great lakes freshwater estuaries: A Cooperative Extension needs assessment model. *Journal of Extension*, 49(1). Retrieved from http://www.joe.org/joe/2011february/a3.php

Shafi, S. M., & Rather, R. A. (2005). Precision and recall of five search engines for retrieval of scholarly information in the field of biotechnology. *Webology*, 2(2). Retrieved from http:/www.webology.org/2005/v2n2/a12.html

Singletary, L., Emm, S., & Hill, G. (2011). An assessment of agriculture and natural resource extension programs on American Indian reservations in Idaho, Nevada, Oregon, and Washington. *Journal of Extension*, 49(2). Retrieved from http://www.joe.org/joe/2011april/a2.php

Sudhakar, P., Poonkuzhali, G., & Kumar, R. K. (2012). Content Based Ranking for Search Engines. International Association of Engineers. Retrieved from http://www.iaeng.org/publication/IMECS2012/IMECS2012_pp497-501.pdf

Tongchim, S., Sornlertlamvanich, V., & Isahara, H. (2006). *Measuring the effectiveness of public search engines on Thai queries*. In: Proceedings of The Fifth IASTED International Conference on communications, internet, and information technology. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.1951&rep=rep1&type=pdf

Walker, E. L., Vaught, C. R., Walker, W. D., & Nusz, S. R. (2011). Attitudinal survey of producers involved in a meat goat artificial insemination clinic. *Journal of Extension*, 49(2). Retrieved from http://www.joe.org/joe/2011april/a6.php

Wyman, M., Escobedo, F., Varela, S., Asuaje, C., Mayer, H., Swisher, N., & Hermansen-Baez. (2011). Analyzing the natural resource Extension needs of Spanish-speakers: A perspective from Florida. *Journal of Extension*, 49(2). Retrieved from http://www.joe.org/joe/2011april/a3.php