

Webology, Volume 2, Number 2, August, 2005

Home	Table of Contents	Titles & Subject Index	Authors Index
----------------------	-----------------------------------	--	-------------------------------

Precision and Recall of Five Search Engines for Retrieval of Scholarly Information in the Field of Biotechnology

[S. M. Shafi](#)

Department of Library and Information Science, University of Kashmir, Srinagar-India
190006

[Rafiq A. Rather](#)

Department of Library and Information Science, University of Kashmir, Srinagar-India
190006

Received May 10, 2005; Accepted August 9, 2005

Abstract

This paper presents the results of a research conducted about five search engines- AltaVista, Google, HotBot, Scirus and Bioweb -for retrieving scholarly information using Biotechnology related search terms. The search engines are evaluated taking the first ten results pertaining to 'scholarly information' for estimation of precision and recall. It shows that Scirus is most comprehensive in retrieving 'scholarly information' followed by Google and HotBot. It also reveals that the search engines (except Bioweb) perform well on structured queries while Bioweb performs better on unstructured queries.

Keywords

Search engine, Precision and recall, Scholarly information, Structured and unstructured queries, World Wide Web

Introduction

The Web is growing as the fastest communication medium. This technology in combination with latest electronic storage devices enable us to keep track of enormous amount of information available to the information society ([Schlichting & Nilsen, 1996](#)). In less than ten years, it has grown from an esoteric system for use by a small community of researchers to the de-facto method of obtaining information for millions of individuals, many of whom have never encountered, and have no interest in the issues of retrieving information from databases ([Oppenheim et al., 2000](#)). A plethora of search engines ranging from general to subject specific are the chief resource discoverers on the Web. These engines search an enormous volume of information at apparently impressive speed but have been the subject of wide criticism for retrieving duplicate, irrelevant and non-scholarly information. The reasons include their comprehensive databases having information on different magnitude like media, marketing, entertainment, advertisement etc. Mainly, these do not sift information from scholar's point of view though some search engines like Google have developed separate applications for disseminating scholarly information like 'Google Scholar' (The tool was incorporated in Google after starting of the

study). The number of search engines that are now available has also made them a popular and an important subject for research ([Clarke & Willet](#), 1997; [Modi](#), 1996).

Related Literature

The growing body of literature on web search engine evaluation is purely descriptive in nature and has little consistency. [Scoville](#) (1996) surveyed a wide range of web search engines for examining the relevance of documents retrievable through them. The first ten hits evaluated for precision have shown Excite, Infoseek and Lycos superior. [Leighton](#) (1996) evaluated the precision of Infoseek, Lycos, WebCrawler and WWWorm using eight reference questions and rated Lycos and Infoseek higher. [Ding and Marchionini](#) (1996) investigated Infoseek, Lycos and Open Text for precision, duplication and degree of overlap using five complex queries. The first twenty hits assessed for precision show that the best results are obtained from Lycos and Open Text. [Leighton and Srivastava](#) (1997) searched fifteen queries on AltaVista, Excite, HotBot, Infoseek and Lycos taking the first twenty hits for evaluation of precision. [Chu and Rosenthal](#) (1996) have investigated AltaVista, Excite and Lycos for their search capabilities and precision. The authors have used ten search queries of varying complexity by evaluating the first ten results for relevance assessment and revealed that AltaVista outperformed Excite and Lycos both in search facilities and retrieval performance. [Clarke and Willett](#) (1997) searched thirty queries of varying nature on AltaVista, Excite and Lycos and obtained best results in terms of precision, recall and coverage from AltaVista. [Bar-Ilan](#) (1998) investigated six search engines using a single query "Erdos". All 6,681 retrieved documents examined for precision, overlap and an estimated recall report that no search engine has high recall.

Objectives

The following objectives are laid down for the study:

- Identification of search engines for retrieval of scholarly information in the field of Biotechnology.
- Assessment of recall and precision of the select search engines.
- Understanding the effect of nature and types of queries on precision and recall of the select search engines.

Method

The process was carried out in three stages. In the first stage, related material available in print and electronic format was collected for the study. In the second stage, search engines were selected and search terms drawn subsequently. In the third stage, the search engines were accessed for the select terms from 25th March to 25th April, 2004. However AltaVista and HotBot were revisited during June 2005 in view of changes in their algorithmic policy. Finally, the data was analyzed for results.

I. Search Engines for the Study

The search engines investigated are:

- AltaVista (General)
- Google (General)
- HotBot (General)
- Scirus (Science & Technology)
- Bioweb (Biotechnology)

II. Sample Search Queries

Twenty search terms were drawn out of a sample of 140 terms compiled with the help of "LC List of Subject Headings" (LCSH, 2003). These were classified under three groups: single, compound and complex terms (Appendix 1) for investigating how search engines control and handle single and phrased terms. Single terms were submitted in natural form, compound terms as suggested by respective search engines and complex terms with suitable Boolean Operators 'AND' and 'OR' between the terms to perform special searches. Five separate queries were constructed for each term in accordance with the syntax of the select search engine.

III. Test Environment

The select search engines offer two modes of searching i.e. simple and advanced mode. The study has chosen the advanced mode of search throughout the study to make use of available features for refining and producing precise number of results. In case of AltaVista and Google "match all of the words" was chosen for single and complex terms and "exact phrase" for compound queries. HotBot and Scirus offer these options through pull down menus. Each search was carried out by choosing title field (i.e. all of the words in title) and limiting age of documents published from 2002 to 2004. All the search engines (except Scirus and Bioweb) were controlled to retrieve the results in English language. Bioweb on the other hand offered relatively different limiting options among which "relevance then date" and hidden Boolean 'OR' were preferred during search.

Each query was submitted to the select engines which retrieved a large number of results but only the first ten results were evaluated to limit the study in view of the fact that most of the users usually look up under the first ten hits of a query. Each query was run on all the five select search engines on the same day in order to avoid variation that may be caused due to system updating (Clarke & Willet, 1997). These first ten hits retrieved for each query were classified as scholarly documents and other categories.

IV. Estimation of Precision and Recall

Precision is the fraction of a search output that is relevant for a particular query. Its calculation, hence, requires knowledge of the relevant and non-relevant hits in the evaluated set of documents (Clarke & Willet, 1997). Thus it is possible to calculate absolute precision of search engines which provide an indication of the relevance of the system. In the context of the present study precision is defined as:

$$\text{Precision} = \frac{\text{Sum of the scores of scholarly documents retrieved by a search engine}}{\text{Total number of results evaluated}}$$

To determine the relevance of each page, a four-point scale was used which enabled us to calculate precision. The criteria employed for the purpose is as under:

- A page representing full text of research paper, seminar/conference proceedings or a patent is given a score of three.
- A page corresponding to an abstract of a research paper, seminar/conference proceedings or a patent is given a score of two.
- A page corresponding to a book or a database is given a score of one.
- A page representing other than the above (i.e. company web pages, dictionaries, encyclopedia, organization, etc.) is given a score of zero.
- A page occurring more than once under different URL is assigned a score of zero.
- A non response of the server for subsequent three searches is assigned a score of zero.

The recall on the other hand is the ability of a retrieval system to obtain all or most of the relevant documents in the collection. Thus it requires knowledge not just of the relevant and retrieved but also those not retrieved ([Clarke & Willet, 1997](#)). There is no proper method of calculating absolute recall of search engines as it is impossible to know the total number of relevant in huge databases. However, [Clark and Willett \(1997\)](#) have adapted the traditional recall measurement for use in the Web environment by giving it a relative flavour. This study also followed the method used by [Clark and Willett](#) by pooling the relevant results (corresponding here to scholarly documents) of individual searches to form the denominator of the calculations. The relative recall value is thus defined as:

$$\text{Relative Recall} = \frac{\text{Total number of scholarly documents retrieved by a search engine}}{\text{Sum of scholarly documents retrieved by all five search engines}}$$

However, in the case of overlapping between search engines results, only the overlapped results are included for the pooling by taking five search engines (say a, b, c, d and e) into consideration which retrieve a₁, b₁, c₁, d₁ and e₁ scholarly documents respectively. Further, where there is no overlap between search engines (i.e. $a \cap b$, $a \cap c$, $a \cap d$ and $a \cap e$ is zero) then the relative recall of search engine 'a' is calculated as $a_1/(a_1+b_1+c_1+d_1+e_1)$. Again if overlapping exists between search engines i.e. $a \cap b = b_2$, $a \cap c = c_2$, $a \cap d = d_2$ and $a \cap e = e_2$ then the relative recall of engine 'a' is $a_1/(a_1+b_2+c_2+d_2+e_2)$. The relative recall is more in case of overlapping between search engines. The mean values for precision and relative recall is obtained by micro-averaging ([Clarke & Willet, 1997](#); [Tague, 1992](#)) i.e. average score for each engine against a query is summed over all the twenty queries and mean value calculated from these totals for single, compound and complex terms separately.

Engines Revisited

Two search engines namely AltaVista and HotBot were revisited during June 2005 to investigate the effect of their changing algorithm policy on precision and recall. The mean precision and recall of the observations in AltaVista show a slight increase while as HotBot shows marginal increase in precision but decrease in its recall value (Table 2).

Results and Discussion

The mean precision and relative recall of select search engines for retrieving scholarly information are presented in Table 1.

Table 1. Mean Precision and Relative Recall of search engines during 2004

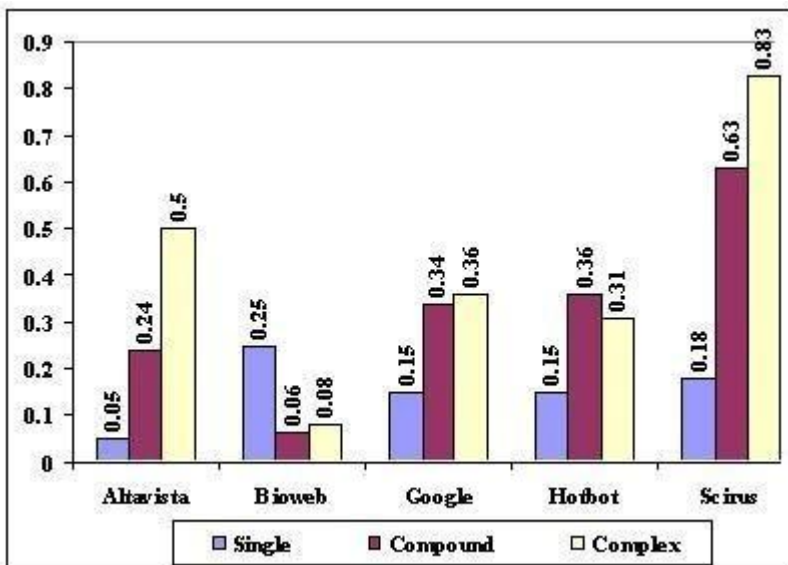
	Altavista	Google	HotBot	Scirus	Bioweb
Precision	0.27	0.29	0.28	0.57	0.14
Recall	0.18	0.20	0.29	0.32	0.05

Table 2. Comparison of mean Precision and mean Recall of AltaVista and HotBot Search engines between 2004 and 2005

Search Engine	Mean Precision 2004	Mean Precision 2005	Mean Recall 2004	Mean Recall 2005
Altavista	0.27	0.29	0.18	0.21
HotBot	0.28	0.33	0.29	0.27

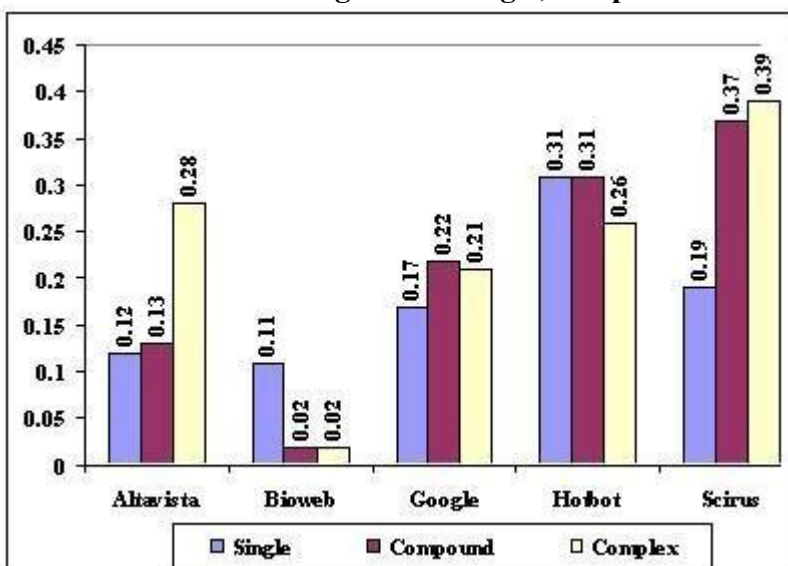
Comparing the mean precision, Scirus scored the highest rank (0.57) followed by Google (0.29) and HotBot (0.28). AltaVista obtained (0.27) while Bioweb received the lowest precision (0.14). The mean precision obtained for single, compound and complex queries of the respective search engines show Scirus as having the highest precision (0.83) for complex queries followed by compound queries (0.63). AltaVista scored the highest precision (0.50) for complex queries followed by compound queries (0.24). Google and HotBot performed better with complex and compound queries while Bioweb performed better with single queries (Figure 1).

Figure 1. Precision of five search engines for single, compound and complex terms



Comparing the corresponding mean relative recall values, Scirus has the highest recall (0.32) followed by HotBot (0.29) and Google (0.20). AltaVista scored a relative recall of 0.18 and Bioweb the least (0.05). While Scirus performed better on complex queries (0.39) followed by compound queries (0.37). HotBot did better in single and compound queries (0.31). Google attained highest recall on compound queries (0.22) followed by complex queries (0.21). AltaVista's performance is better on complex queries (0.28) where as Bioweb performed better on single queries (0.11) (Figure 2).

Figure 2. Relative recall of search engines for single, compound and complex terms



Conclusion

The results depict better performance of Scirus in retrieving scholarly documents and it is the best choice for those who have access to various online journals or databases like Biomednet, Medline plus, etc. Google is the best alternative for getting web-based scholarly documents and its recent introduction of 'Google Scholar' in its beta test for accessing scholarly information offers better dividends for researchers. Scirus acquired the highest recall and precision due to the induction of its journal citations along with web resources; otherwise Google would rank the first. HotBot offers a good combination of recall and precision but has a larger overlap with other search engines which enhance its relative recall over Google search engine. AltaVista once prominent on the Web has lagged behind and the Bioweb is the weakest among the select search engines in all respects. Further, the results reveal that structured queries (i.e. phrased and Boolean) contribute in achieving better precision and recall. The findings also establish the case that precision is inversely proportional to recall i.e. if precision increases recall decreases and vice versa.

References

- 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.
- Chu, H., & Rosenthal, M. (1996). [Search engines for the World Wide Web: a comparative study and evaluation methodology](#). In: *Proceedings of the ASIS 1996 Annual Conference*, October, 33, 127-35. Retrieved August 19, 2003 from <http://www.asis.org/annual-96/ElectronicProceedings/chu.html>
- Clarke, S., & Willett, P. (1997). Estimating the recall performance of search engines. *ASLIB Proceedings*, 49 (7), 184-189.
- Ding, W., & Marchionini, G. (1996). A comparative study of the Web search service performance. In: *Proceedings of the ASIS 1996 Annual Conference*, October, 33, 136-142.
- Leighton, H. (1996, June 25). [Performance of four WWW index services, Lycos, Infoseek, Webcrawler and WWW Worm](#). Retrieved June 10, 2005 from <http://www.winona.edu/library/webind.htm>
- Leighton, H., & Srivastava, J. (1997). [Precision among WWW search services \(search engines\): AltaVista, Excite, HotBot, Infoseek and Lycos](#). Retrieved June 11, 2005 from <http://www.winona.edu/library/webind2.htm>
- Library of Congress (2003). *Library of Congress Subject Headings* (vol.s 1-5). Washington: Library of Congress, Cataloging Distribution Service.
- Modi, G. (1996). Searching the Web for gigabucks. *New Scientist*, 150 (2024), 36-40.
- Oppenheim, C., Moris, A, Mcknight, C., & Lowley, S. (2000). The evaluation of WWW search engines. *Journal of documentation*, 56 (2), 190-211.
- Schlichting, C., & Nilsen, E. (1996). [Signal detection analysis of WWW search engines](#). Retrieved September 15, 2003 from <http://www.microsoft.com/usability/webconf/schlichting/schlichting.htm>
- Scoville, R. (1996). [Find it on the Net](#). *PC World*, January, 14(1), 125-130. Retrieved June 6, 2003 from <http://www.pcworld.com/reprints/lycos.htm>
- Tague, J. (1992). [The Pragmatics of information retrieval experimentation, revisited](#). *Information retrieval experiment*, 14, 59-102. Retrieved 11 June, 2005 from <http://portal.acm.org/citation.cfm?id=149514>

Appendix I Sample Search Queries

- I. Antibiotics
 - II. Biogas
 - III. Brewing
 - IV. Cloning
 - V. Fermentation
 - VI. Gene
 - VII. "enzyme technology"
 - VIII. "gene therapy"
 - IX. "molecular Cloning"
 - X. "monoclonal antibiotics"
 - XI. "protozoa biotechnology"
 - XII. "recombinant DNA"
 - XII. "silage fermentation"
 - XIV. animal AND "genetic engineering"
 - XV +bacterial +starter +cultures
 - XVI. biotechnological AND "process control"
 - XVII. "genetically modified" OR "engineered foods"
 - XVIII. microbial AND "mutational breeding"
 - XIX. "recombinant DNA" AND research
 - XX. "yeast fungi" AND "genetic engineering"
-

Bibliographic information of this paper for citing:

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), Article 12. Available at: <http://www.webology.org/2005/v2n2/a12.html>

[This article has been cited by other articles.](#)

Copyright © 2005, S. M. Shafi & Rafiq A. Rather.