

---

---

## Impact of Query Operators on Web Search Engine Results : An Evaluative Study

Narendra Lahkar

Sanjib K Deka

### Abstract

*Study on Web search services have reported that about 90% of Web searchers using simple queries and only 10% use advanced query operators. In this study, three search engines, Google, Yahoo and MSN, are compared in the areas of coverage and duplication in terms of use of Query Operators. Google is found at the top in terms of coverage, as compared with the other two search engines, i.e. MSN Search and Yahoo!. Analysis shows that there is a significant change in coverage when using query operators in all the three search engines. It is found that correct use of query operators would increase the effectiveness of web searching.*

**Keywords :** Search Engines, Search Strategies, Search Techniques.

### 0. Introduction

The World Wide Web (WWW) is a very large distributed digital information space. It is an extensive hypertext system connecting a worldwide network of information nodes, each is located on one or more server machines. Some estimates suggest that the Web currently includes about 150 million pages and that this number doubles every four months. The ability to search and retrieve information from the Web efficiently and effectively is an enabling technology for realizing its full potential. A variety of systems have been developed to provide effective access to these resources. A Web browser, such as Netscape or Internet Explorer, provides an interface to the Web that runs on a client machine. Web search engines, are information retrieval systems that locate nodes in response to queries. Search engines generally provide a variety of search options ranging, from simple lists of keywords to more elaborate queries including Boolean and proximity operators, truncation, field restrictions, and phrase matching.

Current Information Retrieval (IR) techniques are not knowledge-enabled and hence cannot give precise answers to precise questions. To overcome this problem, a current trend on the Web is to let users annotate documents using metadata languages. Conventional wisdom holds that more precisely and narrowly formulated Internet queries not only are much less likely to result in overwhelming numbers of hits but also to give better results. Research has reported that about 10% of Web searchers utilize advanced query operators, with the other 90% using extremely simple queries.<sup>4</sup> It is often assumed that the use of query operators, such as Boolean operators and phrase searching, improves the effectiveness of Web searching. Searchers seldom use advanced query structure, such as Boolean operators or phrase searching, when using IR systems. This behavior has been especially characteristic of Web searchers. The usage of Boolean operators is typically about 10%. It is often assumed that correct use of query operators would increase the effectiveness of Web searches. Advanced query operators are well known, and many of these techniques (e.g., PHRASE searching and MUST APPEAR operators) are generally easy to employ and available. Most present generation Web search engines support a variety of Boolean and other query operators and often recommend them as a way to improve searches. However, based on a review of the published research on Web searching, it appears that the majority of Web searchers continue to use very simple queries, with little or no use of Boolean or other query operators. Additionally, Web search engines continue to attract large numbers of Web searchers. Many of the most popular Web sites in terms of number of visitors are Web search engines, implying that most users view search engines as the best method available for finding information on the Web.<sup>4</sup> There are millions of Web users and about

---

85% of them use search engines to locate information on the Web.<sup>1</sup> It is determined that search engine use is the second most popular Internet activity next to e-mail. Nearly everyone is familiar with the experience of searching with a Web search engine and using a search interface to search a particular Web site. Due to high demand there are hundreds of general purpose and thousand of specialized search engines.

Are the Operators a vital aspect of effective Web searching? Are they useful for most searchers and most searches? Do they in fact provide more accurate, precise or focused searches? In this article, it has been tried to investigate these questions and to provide a partial answer. The specific objective of this study is to analyses the effect of query operators on the coverage, i.e. the number of results found, by Web search engines in response to submitted queries. In this study, four queries were used to search on three selected Search Engines, i.e. Google, Yahoo and MSN Search, using a variety of formulations.

## **1. Search Engines**

Automated methods for retrieving information on the Web can be broadly classed as search tools or search services. Search tools employ robots for indexing Web documents. They feature a user interface for specifying queries and browsing the results. At the heart of a search tool is the Search Engine, which is responsible for searching the index to retrieve documents relevant to a user query. Search services provide users a layer of abstraction over several search tools and databases and aim at simplifying the Web search. People use search engines for finding information on the web. A search engines is an information retrieval systems, which is used to locate the web pages relevant to user queries. A web search engine contains indexing, storage, query processing, spider and user interface subsystems.

## **2. Query Operators**

A query to a web search engine usually consists of a list of keywords, to which the search engine responds the Internet. This makes users difficult to find their desired information easily. Search engines provide a way for users to locate web information that satisfy their requirements. Users can specify their information needs with multiple keywords that should appear in the returned documents. In order to define the query more accurately, many search engines allow mathematical operators like '+', '-', quotation marks, brackets, Boolean operators like, AND, OR, NOT or even more advanced operators like 'title:', 'url:', etc. in the query. These operators define the co-occurrence and exclusion of terms, the proximity between terms, and even the exact position a term should appear in a document. With these operators, more exact queries can be defined. This is must to use these query operators in searching the web as to minimized search time, to get exact required information and to overcome unnecessary junk hits.<sup>1</sup>

## **3. Methodology**

With this study, it has been tried to investigate the effect of using queries with operators on the results retrieved by Web search services relative to the results retrieved by queries with no operators. In the present study, selected queries were searched on three most used search engines; Google, Yahoo and MSN Search, on the same day. All queries were searched using the same web browser, i.e. Opera. The following question was set for this study, Will there be a change in coverage (i.e., the number of results found) when using query operators relative to no query operators ?

### **3.1 Selection of Query Terms**

The process of measuring retrieval effectiveness requires users queries. The query terms for this study have been selected from a survey carried out amongst the research scholars of IIT Guwahati in the year 2003. Four randomly selected broad subjects have been used as query terms for the study. These are,

---

PHYSICS CONDENSED MATTER, ORGANIC CHEMISTRY OF NATURAL PRODUCTS, WATER RESOURCES RESEARCH, and DIGITAL SIGNAL PROCESSING. As a technical institute library professional, I have confronted with users with all these query terms in different perspective. Keeping this on mind, the terms have been selected. We then generated the following query logs with three query operators. For this study three query operators are used, these are, AND, MUST APPEAR (+), and PHRASE (""), because these operators are applicable with all the three selected search engines; Google, Yahoo and MSN Search.

**With AND**

Digital AND Signal AND Processing  
Organic AND Chemistry AND Natural Products  
Physics AND Condensed AND Matter  
Water AND Resources AND Research

**With MUST APPEAR (+)**

+Digital +Signal +Processing  
+Organic +Chemistry +of +Natural +Products  
+Physics +Condensed +Matter  
+Water +Resources +Research

**With PHRASE ( " )**

"Digital Signal Processing"  
"Organic Chemistry of Natural Products"  
"Physics Condensed Matter"  
"Water Resources Research"

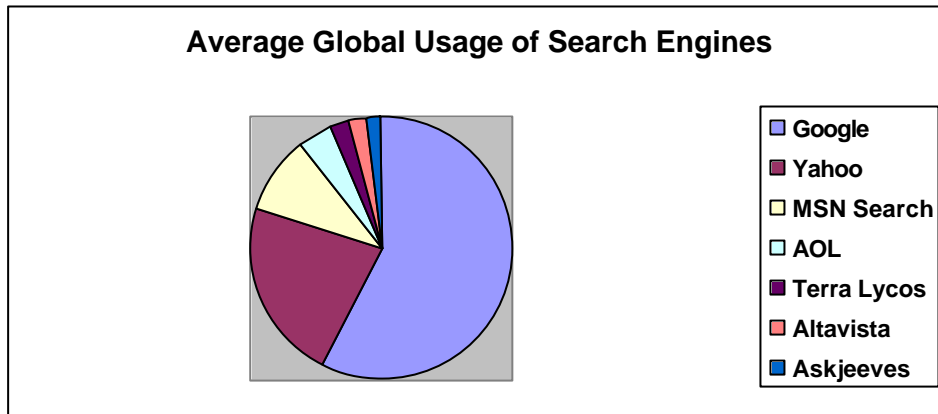
### 3.2 Selection of Documents

The next issue was what number of documents to compare. Studies show that approximately 80% of Web searchers never view more than the first ten results in a results list<sup>4</sup> Based on this overwhelming evidence of Web searcher behaviour, we utilized only the first ten results in the results list for comparison of coverage and duplication in this study. For the analysis of coverage, we utilize the reported number of documents by the respective search engines.

### 3.3 Searching Environment

The next step was to select what Web search services to utilize. Search engines are the major portals for users of the Web, with 71% of Web users accessing search engines to locate other Web sites. There are approximately 3,200 search engines on the Web, with a handful dominating in terms of usage. We have selected three most used Search Engines for this study, which are Google, Yahoo and MSN. OneStat.com (www.onestat.com), the number one provider of real-time intelligence web analytics has reported on May 2003, that Google is the most popular search engine on the Web, having global average usage share of 55.2 percent during the survey period.<sup>9</sup> According to their survey, the following seven are the largest search engines on the Web:

- |                        |                       |
|------------------------|-----------------------|
| 1. Google (55.2 %)     | 2. Yahoo (21.7%)      |
| 3. MSN Search (9.6 %)  | 4. AOL Search (3.8 %) |
| 5. Terra Lycos (2.6 %) | 6. Altavista (2.2 %)  |
| 7. Askjeeves (1.5 %)   |                       |



### 3.4 Data Collection Method

In this study, the four original queries were submitted to one of the search engines. Then the queries were modified by adding the advanced searching operators and submitted it to the same search engine. We repeated the process for all four queries and all search engines. After submitting each query, the number of reported retrieved documents and the retrieving time were recorded, wherever available. Our primary focus was to evaluate the effect, if any, of query operators on a particular search engine for a query. We compared the results from the original queries to the results of the modified queries with operators on each search engine in the areas of coverage and duplication.

## 4. Discussion of Results

On the basis of the survey, we report here the result in the areas of coverage. We present here the descriptive statistics regarding the use of operators and search terms.

### 4.1 Coverage

There were over 42,484,458 documents retrieved by all the queries on all three search engines. The maximum coverage was 3,120,000 and the minimum coverage was 116. The analysis results presented in the following table shows that the PHRASE operator resulted in a decrease in coverage across all three search engines in all the queries. The MUST APPEAR operator resulted in an increase in coverage on YAHOO, but it decreases on MSN. This increase in coverage on MSN is somewhat surprising as usually MUST APPEAR operator is thought of as an operator to decrease the number of retrieve results. The AND operator resulted in decrease in coverage on GOOGLE and YAHOO, but no change on MSN, except with third query. The AND operator is also typically used to decrease the number of retrieve results. Although AND and MUST APPEAR operators are conventionally used to narrow a query, but they did not do so in our analysis. It is interesting to note that, despite the fact that AND is the default operator on GOOGLE, searches using AND and searches using with no operators did not return identical results lists, even when repeated. Our results indicate that except PHRASE operators, there is no significant change in coverage incase of other two operators, as compared to the coverage with no operator.

**First Query : Condensed Matter Physics**

Query Type	Google	Yahoo	MSN
WITHOUT OPERATOR	534,000	506,000	67,076
MUST APPEAR (+)	549,000	515,000	66,846
AND	531,000	505,000	67,076
PHRASE ("")	158,000	194,000	24,192

**Second Query : Digital Signal Processing**

Query Type	Google	Yahoo	MSN
WITHOUT OPERATOR	2,450,000	1,960,000	402,932
MUST APPEAR (+)	2,440,000	2,000,000	396,236
AND	2,440,000	1,950,000	402,932
PHRASE ("")	565,000	504,000	103,168

**Third Query : Organic Chemistry of Natural Products**

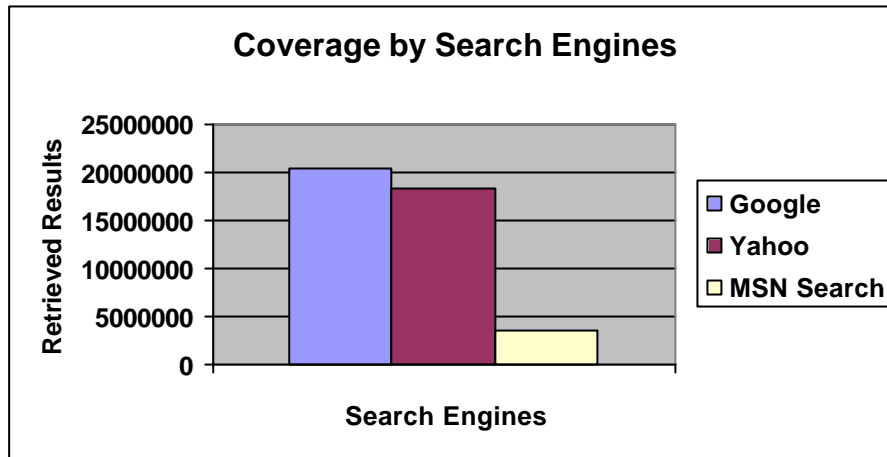
Query Type	Google	Yahoo	MSN
WITHOUT OPERATOR	552,000	389,000	72,478
MUST APPEAR (+)	443,000	390,000	72,473
AND	550,000	385,000	71,500
PHRASE ("")	166	316	116

**Forth Query : Water Resources Research**

Query Type	Google	Yahoo	MSN
WITHOUT OPERATOR	3,120,000	3,000,000	613,661
MUST APPEAR (+)	3,060,000	3,070,000	613,661
AND	3,040,000	3,010,000	613,661
PHRASE ("")	38,100	39,600	8,268

**4.1.1 Coverage by Search Engine**

Measuring the coverage effect of query operators on a particular search engines is a subtle task. Using the above set of queries, we found that there was a significant difference in the coverage on all three search engines. This is mainly because of the size of the databases, algorithms applied for indexing and difference in counting techniques. GOOGLE retrieved far more documents than the other two search engines. GOOGLE retrieved 20,470,266 documents, YAHOO retrieved 18,417,916 documents and MSN retrieved the less number of documents i.e. 3,596,276. It has also been noticed that only GOOGLE has resulted duplicate hits in all searches among the first ten hits.



## 5. Conclusions And Suggestions

Both search engine and query operator has a significant effect on coverage. Of course, these hits correspond to the matching documents identified by the system. It is hard to pin down the exact number of documents indexed by search engines because of web volatility and different counting and estimation techniques. Most current Internet users prefer to use relatively short and simple queries and have trouble using more complex search features even when they try. Since it may well be the case that effective performance is not improved even if more complex queries are constructed, there may be little point in exhorting users to make use of these search features.<sup>3</sup>

Web searching involves not only selection of terms, but also the construction of queries and selection of search engines. So, there needs more research from the perspectives of both the search engine designer and the search engine user. In this study, we focused mainly on the perspective of the user. The most important avenues of further research to pursue are those that would provide further insight into the issues of relevance and ranking, since these are of more concern to most users than coverage. Certainly, results indicate that research in the improvement of Web and perhaps other IR systems should focus on areas beyond Boolean and other query operators.

There is one important query requirement that is currently not being covered by these search engines. A user often wants to find information about a topic that is in the context of another topic. According to current search engines, however, there is no query operator to define this relation of context inclusion between query terms. The closest is the '+' operator which only specifies two terms to appear simultaneously in a document. This might lead to user's constraint to express their need accurately. In order to implement the task of finding one topic in the context of the other in one pass, web search engine needs to introduce a new search operator 'in'. This operator defines that the searching of one query term should be in the context of the other query term.

Indexing quality has an overwhelming effect on retrieval effectiveness. It has been called one of the grand challenges in the digital libraries realm. Comprehensively indexing the entire Web and building one huge integrated index will only further deteriorate retrieval effectiveness, since the Web is growing at an exponential rate. Unless the current generation of search tools and services significantly improve their retrieval effectiveness, the Web will continue to evolve toward an information entertainment center for users with no specific search objectives.<sup>5</sup>

## 6. References

1. CAN, FAZLI and OTHERS, "Automatic Performance Evaluation of Web Search Engines", Information Processing and Management, v.40, 2004, pp.495-514 5
2. CHI1, CHI-HUNG, DING, CHEN and LAM, KWOK-YAN, "Context Query in Information Retrieval", Proceedings of the 14th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'02), 2002 IEEE 4
3. EASTMAN, CAROLINE M. "30,000 Hits May Be Better Than 300: Precision Anomalies in Internet Searches", ACM SIGIR, 99 Berkley, USA, pp. 13-14 3
4. EASTMAN, CAROLINE M. and JANSEN, BERNARD J. "Coverage, Relevance, and Ranking: the Impact of Query Operators on Web Search Engine Results", ACM Transactions on Information Systems, v. 21(4), 2003. pp 383-411. 1
5. GUDIVADA, VENKAT N. and RAGHAVAN, VIJAY V. "Information Retrieval on World Wide Web", Ieee Internet Computing, no.4, 1997, pp. 58-68. 2
6. JANSEN, B J and POOCH, U, "A Review of Web Searching and a Framework for Future Research." Journal of the American Society for Information Science and Technology, v.25 (3), 2001, pp.235-246. 6
7. KOBAYASHI, M and TAKEDA, K, "Information Retrieval on the Web", ACM Computing Surveys, v.32 (2), 2000, pp.107-109 7
8. LUPINACCI, J A, "How To Find Useful Information on the Internet", <www.techlearning.com> Access on 07-04-2004 8
9. Search Engine Ratings. < www.onestat.com > Access on 20-01-2004 9
10. SULLIVAN, DANNY, "The Major Search Engines And Directories", <www.searchenginewatch.com> Access on 28-04-2004 10

## About Authors



**Dr. Narendra Lahkar** is the Head of the Department of Library and Information Science, Guwahati University. He has done his MLISc from Guwahati University, M.Sc. in Information Studies from University of Sheffield (UK) and Ph.D. from Guwahati University. He is the Chairman of the sub-committee for drafting the Library Legislation of Assam. He is President, Assam Library Association, Zonal Secretary, Indian Association of Teachers in Library and Information Science (IATLIS). Besides attending number of international, national and regional conferences, he has contributed professional papers also. He has edited two publications.



**Mr. Sanjib K Deka** is working as Technical Assistant IIT, Guwahati since 1999. He holds M.LISc & PG Diploma in Computer Applications. He is life member of Guwahati Library Association. He has presented paper on Information needs in Science & Technology for Social & Technology for Social Development during 12th Conference of Assam Library Association.