Coverage, Relevance, and Ranking: The Impact of Query Operators on Web Search Engine Results

CAROLINE M. EASTMAN
University of South Carolina
and
BERNARD J. JANSEN
The Pennsylvania State University

Research has reported that about 10% of Web searchers utilize advanced query operators, with the other 90% using extremely simple queries. It is often assumed that the use of query operators, such as Boolean operators and phrase searching, improves the effectiveness of Web searching. We test this assumption by examining the effects of query operators on the performance of three major Web search engines. We selected one hundred queries from the transaction log of a Web search service. Each of these original queries contained query operators such as AND, OR, MUST APPEAR (+), or PHRASE (" "). We then removed the operators from these one hundred advanced queries. We submitted both the original and modified queries to three major Web search engines; a total of 600 queries were submitted and 5,748 documents evaluated. We compared the results from the original queries with the operators to the results from the modified queries without the operators. We examined the results for changes in coverage, relative precision, and ranking of relevant documents. The use of most query operators had no significant effect on coverage, relative precision, or ranking, although the effect varied depending on the search engine. We discuss implications for the effectiveness of searching techniques as currently taught, for future information retrieval system design, and for future research.

Categories and Subject Descriptors: H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*query formulation*, *search process*; H.3.4 [**Information Storage and Retrieval**]: Systems and software—*performance evaluation* (*efficiency and effectiveness*); H.3.5 [**Information Storage and Retrieval**]: Online Information Services—*web-based services*

General Terms: Human Factors, Performance, Experimentation

Additional Key Words and Phrases: Relative precision, coverage, ranking, Boolean operators, query operators, search engines, Web results

Authors' addresses: C. M. Eastman, Department of Computer Science and Engineering, University of South Carolina, Columbia, SC 29208; email: eastman@cse.sc.edu; B. J. Jansen, School of Information Sciences and Technology, The Pennsylvania State University, University Park, PA 16801; email: jjansen@acm.org.

1. INTRODUCTION

Searchers seldom use advanced query structure, such as Boolean operators or phrase searching, when using information retrieval (IR) systems [Borgman 1996]. This behavior has been especially characteristic of Web searchers. Numerous publications of Web research have noted the near absence of query operators in Web queries [Hölscher and Strube 2000; Jansen et al. 2000b; Spink et al. 2002]. 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 by increasing the total number of retrieved documents, increasing the number of relevant documents retrieved, or improving the ranking of relevant documents. For example, many university library Web sites offer online explanations on the use of Boolean and other query operators. Many online digital libraries offer Boolean searching, such as the ACM Digital Library (http://portal.acm.org/portal.cfm) and the IEEE Xplore (http://ieeexplore.ieee.org).

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. If training is needed, one can find an abundant number of articles and books on advanced searching strategies (e.g., Korfhage [1997]) and numerous educational courses on searching strategies. For Web users, there are online tutorials specifically targeted for Web searching [Sullivan 2000, 2002]. The use of Boolean searching is also common in online databases, such as LexisNexis [2003]. Most major Web search engines support a variety of Boolean and other query operators [Notess 2003] and often recommend them as a way to improve searches.

Google's [2003] advice on advanced search states "You can increase the accuracy of your searches by adding operators that fine-tune your keywords." America Online Search's (AOL) advice page [AOL 2003] states, "There are times when you might want the precise results that a Boolean query provides." Microsoft Search's (MSN) help page [MSN 2003] advises that "An advanced search has more options than a basic search." and that one should "Use the Boolean operators AND, OR, and NOT to focus your search." Although statements such as these clearly do not indicate that the use of advanced search operators is required for effective searching, they do indicate that they are expected, in general, to give improved results.

There has also been significant research on the development of advanced searching features for IR systems. Active research using Boolean queries and systems is ongoing within the areas of information systems development [Chang et al. 1999; Clarke and Cormack 2000], query modeling [Hiemstra and Vries 2000], system evaluation [Sormunen 2000], education [Ford et al. 2003] and information science [Lucas and Topi 2002]. The effect of query formulation on system effectiveness continues, although most research has been on traditional IR systems [Chowdhury et al. 2002]. Frants and colleagues [1999] provide a historical view of Boolean IR, highlighting the stream of research on Boolean systems.

Based on the volume of literature on the subject of Boolean searching, one might think it was an extremely common search tactic for information seekers. However, based on a review of the published research on Web searching [Jansen and Pooch 2001; Spink et al. 2002], it appears that the vast majority of Web searchers continue to use very simple queries, with little or no use of Boolean or other query operators.

Why aren't Web searchers using more advanced queries? Some researchers have stated that Web searchers are lazy [Zapur and Zhang 2000] or that there is a design flaw with Web search engines [Clark 2001]. Neither assertion seems likely. There are studies and data that suggest Web searchers may be finding the information they want using these simple queries. In a survey of Web searchers, approximately 70% of the searchers stated that they had located relevant information on the search engine [Spink et al. 1999]. 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 [Cyber Atlas 2002], implying that most users view search engines as the best method available for finding information on the Web.

Is the support of Boolean and other advanced operators by search engines an anachronism? Or are they a vital aspect of effective Web searching? Are these operators useful for some searchers and some searches? Are they useful for most searchers and most searches? Do they in fact provide more *accurate*, *precise* or *focused* searches? In this article, we investigate these questions and provide a partial answer.

The specific objective of this study is to determine the effect of query operators on the number, relevance and ranking of results retrieved by Web search engines in response to queries submitted by a general population of Web search engine users. This knowledge is essential to understanding user-searching behavior on the Web, for the development of instructional materials for Web searching, and for improving the design of future IR systems. In this paper, we present a review of literature, our research methodology, and the research results from the perspectives of coverage, relative precision, and ranking. We also discuss the results and the directions for future research.

2. RELATED STUDIES

There have been relatively few studies comparing the retrieval results of different search engines using various query formulations [Eastman 2002; Gudivada et al. 1997; Jansen 2000; Lucas and Topi 2002; Petersen 1997; Selberg and Etzioni 2000]. Petersen [1997] presents comparisons using unrelated queries (i.e., "embargo" as a single search term and "Woodrow Wilson's Fourteen Points" as a phrase). This type of comparison provides limited insight on the impact of query operators for the same topic. Using two queries and multiple search engines, Gudivada et al. [1997] present data addressing coverage (i.e., number of documents retrieved). Boolean and phrase queries are compared, showing that the AND queries generally retrieved fewer hits than the OR version. For these two queries, the phrase queries retrieved fewer results than the AND queries.

In the course of reviewing the changes in search engine results over time, Selberg and Etzioni note that there was a smaller percentage change over time in the results retrieved by the PHRASE and MUST APPEAR operators when compared to the default query with no operators [Selberg and Etzioni 2000]. Jansen examines the similarity in results for queries submitted to five search engines using different searching operators, reporting a 70% similarity in results between the queries with no operators and the queries with operators [Jansen 2000]. Lucas and Topi use eight search topics from which naïve and expert queries were formulated and submitted to various Web search engines to evaluate relevancy [Lucas and Topi 2002]. Terms could change between the naïve and expert queries, and the query retrieving the most relevant results was the expert query. The researchers report that term selection and incorrectly formulated queries were the primary cause for most of the relevancy variation. Eastman [2002] explores the precision of search engines using a variety of topics and query formulations, noting that precision did not necessarily improve with the use of the advanced query operators.

Other studies examine the difference between average and more sophisticated Web queries [Hölscher and Strube 2000; Silverstein et al. 1999; Spink et al. 2002]. Some studies separate Boolean usage from the use of other query operators. Spink et al. [2002] show a Boolean usage rate of about 10% for Excite users. Jansen et. al. [2002b] report a Boolean usage of 8.5% and a usage of other query operators at approximately 15.5%. Other studies report the combined use of Boolean with other operators. Silverstein et al. [1999] report an advanced operator usage of approximately 20% for Alta Vista users. In a study of novice and expert users of Fireball, a European-based search engine, Hölscher and Strube [2000] report greater use of the advanced searching options for the experts. None of these studies examined the effect of the query operators on Web search results.

We could locate no study researching the impact of query operators from a large number of queries from real users across multiple search engines that controlled for operators and term changes. Controlling for query terms is important in such studies as research shows that term selection can dramatically alter retrieved results [Spink 1995].

3. RESEARCH DESIGN AND METHODOLOGY

We 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, controlling for other factors such as term usage, IR system, and document collection. This section describes the specific research questions and the methodology used to investigate them.

3.1 Research Questions

Our research questions are:

(1) Will there be a change in coverage (i.e., the number of results found) when using query operators relative to no query operators?

Hypothesis 1a. The use of the AND query operator will result in a decrease in coverage.

Hypothesis 1b. The use of the OR query operator will result in an increase in coverage.

Hypothesis 1c. The use of the MUST APPEAR query operator will result in a decrease in coverage.

Hypothesis 1d. The use of the PHRASE query operator will result in a decrease in coverage.

Since a searcher may want to increase coverage or decrease coverage based on the information need, we are interested in either an increase or a decrease in coverage, depending on the operator utilized. The AND, MUST APPEAR, and PHRASE operators are generally utilized to narrow a query, while the OR operator is generally utilized to broaden a query. Typically, one implements the MUST APPEAR operator by placing a + immediately preceding a query term. One usually implements the PHRASE operator by placing two or more terms within " ".

(2) Will there be an increase in relative precision when using query operators relative to no query operators?

Hypothesis 2a. The use of the AND query operator will result in an increase in relative precision.

Hypothesis 2b. The use of the OR query operator will result in an increase in relative precision.

Hypothesis 2c. The use of the MUST APPEAR query operator will result in an increase in relative precision.

Hypothesis 2d. The use of the PHRASE query operator will result in an increase in relative precision.

For the second research question, we are assuming that searchers will generally want to increase relative precision, rather than just change it. We define relative precision as the number of relevant documents within the top ten ranked and retrieved documents. If there are fewer than ten results, then we use that number to determine relative precision.

(3) Will relevant documents be ranked higher when using query operators relative to not using query operators?

Hypothesis 3a. The use of the AND query operator will result in relevant documents being ranked higher.

Hypothesis 3b. The use of the OR query operator will result in relevant documents being ranked higher.

Hypothesis 3c. The use of MUST APPEAR query operator will result in relevant documents being ranked higher.

Hypothesis 3d. The use of PHRASE query operator will result in relevant documents being ranked higher.

For the third research question, we are interested in whether the use of query operators causes relevant documents to be ranked higher, rather than

just causing a change in ranking. We did not track whether or not an individual document changed in rank, but whether or not there was any relevant document at a particular given rank. Similar to research question number two, most searchers desire relevant documents higher in the results list. Even though most users only examine the first ten or so documents, it is usually viewed as desirable to have the relevant documents ranked higher [Cooper 1968].

4. METHODOLOGY

The methodology of our research is outlined in the following sections.

4.1 Selection of Queries

Our first step was query selection. Research shows that about 10%–20% of Web queries contain query operators [Jansen et al. 1998; Silverstein et al. 1999]. Research of Web searching also indicates that Web queries cover a variety of topics [Wolfram 1999] and are primarily noun phrases [Jansen et al. 2000a; Kirsch 1998].

We selected the specific queries used in this research from a transaction log of a subset of queries submitted to the Excite search service on 1 May 2001. Excite supports several advanced query operators. The operators used in this study are the AND, OR, MUST APPEAR, and PHRASE searching operators.

We eliminated all queries that did not contain one of these operators from the transaction log. We also eliminated all queries that were obviously seeking pornography, as determined by the researchers. We then generated four transaction logs, one for each of the query operators used in this study. We then qualitatively reviewed each of the queries in the four transaction logs, removing those queries that were improperly constructed (e.g., AND is misspelled, the trailing quote is missing on a PHRASE search, the Boolean operator is the last term in the query, etc.). Since we were not investigating the effectiveness of improperly formed queries, we did not want these queries to skew our results. The queries using more than one distinct operator were also removed. Of the remaining queries in each transaction log, we randomly selected 25 from each transaction log for use in this study.

Twenty-five of the queries selected contained the AND operator; twenty-five contained the OR operator; twenty-five contained the MUST APPEAR operator; and twenty-five contained the PHRASE operator. Each query contained one or more uses of the same operator. Query lengths ranged from two to eight terms. We did not count the Boolean and other operators as terms. Appendix A contains the entire set of queries used for this research.

4.2 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 [Hölscher and Strube 2000; Jansen et al. 2000b; Silverstein et al. 1999]. Based on this overwhelming evidence of Web searcher behavior, we utilized only the first ten results in the results list for comparison of relative precision and ranking in this study. If duplicates occurred within the first ten,

we used only one of the duplicates in the analysis of relative precision and ranking of relative documents. For the analysis of coverage, we utilize the reported number of documents by the respective search engines. These reported numbers might contain duplicates, and they depend on the estimating algorithms used by the respective search engines. The strength of this research is that we are utilizing real search engines with real document collections. This strength is also a limitation in that previous research has shown that these document collections may change over time [Selberg and Etzioni 2000].

4.3 Searching Environment

We then had 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 [NielsenMedia 1997]. There are approximately 3,200 search engines on the Web [Sullivan 2000], with a handful dominating in terms of usage. These include AOL, Google and MSN, which are the search engines used for this research. Our selection criterion was that these are the three most popular Web search engines in terms of number of unique visitors per month. Yahoo!, also an extremely popular Web information service, is a Web directory and also utilized a third party for its search services at the time of the study. We provide a brief overview of the three search services utilized in this research.

AOL is America Online's search engine. Since August of 1999, AOL has utilized third parties for its backend document collection, first using Excite, then Inktomi, and now Google. The searches for this research were carried out during the period when AOL used Inktomi. AOL currently has access to over 2 billion documents from the Google database. AOL received 90,031,000 unique visitors in August 2002 [Nielsen Netrating 2002].

Google is a full-featured Web searching tool. In addition to possessing a searchable database of over 2 billion HTML documents, Google has indexed 700,000,000 USENET messages, 35,000,000 non-HTML documents, and 390,000,000 images. Google reports approximately 150 million search queries per day. In March 2002, Google received 31,901,000 unique visitors [Nielsen Netrating 2002].

MSN uses the Inktomi Gigadoc database and the LookSmart directory service as its backend content collection. LookSmart reports its service indexes over 2,500,000 unique uniform resource locators (URLs) in 250,000 categories [LookSmart 2003]. Inktomi reports having a master database containing over two billion URLs [Inktomi 2002]. In March 2002, MSN received 97,426,000 unique visitors [Nielsen Netrating 2002].

4.4 Searching Rules

All the search engines supported all the query operators in some form at the time of the study from their main page. At the time of the study, AOL directly supported the use of the AND, OR, MUST APPEAR, and PHRASE searching, although it also provided an advanced search option which facilitated the use of operator functionality. Google directly supported the AND, OR, MUST

APPEAR, and PHRASE searching, although it states that the use of AND is not necessary. MSN directly supported the AND, OR, and MUST APPEAR operator. There was a drop down box for PHRASE searching. All search engines also provided an advanced search mode, which directly supported all of the operators considered here as well as other operators and features, such as field restrictions.

5. DATA COLLECTION METHOD

We submitted each of the 100 original queries to one of the search engines. We then modified the query by removing the advanced searching operator(s) and submitted it to the same search engine. As examples, queries with the MUST APPEAR operator (+*gale* + *research* +*company*), the PHRASE operator ("*gale research company*"), the AND operator (*gale* AND *research* AND *company*), or the OR operator (*gale* OR *research* OR *company*) would be modified to the query (*gale research company*). The entire process of submitting the original and modified query pair took approximately five minutes or less on each search engine. Therefore, the opportunity for the document collection to change between query submissions was minimal. We repeated the process for all queries and all search engines. Data collection occurred from March 12 to March 22, 2002.

After we submitted each query, we recorded the number of reported retrieved documents. Additionally, we saved the URLs for the top ten results. We did not evaluate identifiable Sponsored Links or Sponsored Sites. Typically, search engines present these sponsored links or sites within a separate area of the results page, usually on the righthand side of the browser window. MSN lists Featured and Sponsored sites at the top of the results listing, under separate headings. MSN also lists those results from its Web Directory sites.

Four independent reviewers who had not performed the original searches evaluated each of the retrieved Web sites. They first reviewed the summary presented by the search engine; they then retrieved the full text of the Web site if the summary did not provide sufficient information for evaluation. Each of the four reviewers made independent relevance judgments on each of the sites based on the reviewer's interpretation of the original query terms using topical relevance [Hawking 2000]. Reviewers evaluated the results of each search separately. We provided each of the reviewers a written explanation of the reviewing and the relevance judgment tasks. The written instructions are included in Appendix B. There are other forms of searching where the evaluation criteria might be different, including known item, home page, name, Goggle Whacking, and vanity searching.

The reviewers rated each document using a four-point relevance scale similar to instruments utilized in other studies [Spink et al. 1998b]. A score of 4 indicated a totally relevant document. A score of 3 indicated a partially relevant document. A score of 2 indicated a somewhat relevant document. A score of 1 indicated a nonrelevant document. The average relevance score had to be 3.00 or higher to be deemed as relevant for this study. The calculated agreement across the four raters using the individual reviewer's original rating for each document (i.e., 1, 2, 3, or 4) was quite reasonable (0.81). The reviewers were

Table I. Paired Sample t-test with and without Operators for Coverage across Three Search Engines

Operator	AOL	Google	MSN
AND	1.04 (df = 24)	1.62 (df = 24)	**2.46* (df = 24)**
Mean Without	284,607	366,964	57,197
Mean With	290,310	383,154	483,133
OR	1.71 (df = 23)	**3.02** (df = 23)**	0.77 (df = 24)
Mean Without	56,105	198,095	6,970,033
Mean With	59,138	1,364,270	10,031,680
MUST APPEAR	−2.01 (df = 22)	−1.00 (df = 24)	**2.89** (df = 22)**
Mean Without	124,999	254,876	6,809
Mean With	109,000	254,766	138,521
PHRASE	**−3.95** (df = 23)**	**−9.13** (df = 24)**	**−5.73** (df = 24)**
Mean Without	94,812	227,572	34,498
Mean With	67,838	5,047	1,118

Notes: (1) $^{**}p < 0.01$, (2) $^{*}p < 0.05$, (3) For ease of viewing, we present actual coverage means rather than the log values.

aware of the identities of the search engines. Our primary focus was to evaluate the effect, if any, of query operators on a particular search engine. Given the independent searches, four independent reviewers conducting separate evaluations, and the high inter-rater reliability score, we consider the evaluation reasonably unbiased, especially with respect to the impact of operator usage.

For each of the three areas (coverage, relative precision and ranking), we compared the results from the original queries with each operator to the results of the modified queries without operators on each search engine using a paired sample t-test. The paired sample t-test determines whether or not means are distinct between samples (i.e., two-sided) and does not assume that the variances are equal.

In the next section, we present the empirical results from our analysis.

6. RESULTS

We report the results in the areas of coverage, ranking, and relevance. We present the descriptive statistics regarding the usage of operators and search terms first, with the discussion following in the next section.

6.1 Coverage

There were over 560,000,000 documents retrieved by all the queries on all three search engines. The average coverage was 964,374 documents with a standard deviation of 5,568,384. The maximum coverage was 61,540,009. The minimum coverage was 0. Given the size and distribution of the coverage means, we utilized the log of the coverage for the t-test analysis.

Our first research question and associated hypotheses were that the use of query operators would result in a change in coverage. Table I presents the values of the paired t-tests along with the degrees of freedom (df) and actual means.

The analysis results presented in Table I shows that the PHRASE operator resulted in a decrease in coverage across all three search engines. The MUST

Table II. Results of Pairwise Comparisons of Coverage by Search Engine

Search Engine Comparison	Difference Between Coverage Means	Simultaneous 95% Confidence Limits
MSN—Google	**1,944,761***	590,429–3,299,093
MSN—AOL	**2,175,217***	801,112–3,549,321
Google—AOL	230,456	−1,143,649–1,604,560

Note: (1) $^*p < 0.05$.

APPEAR operator resulted in an increase in coverage on MSN but no change in coverage on AOL or Google. This increase in coverage on MSN is somewhat surprising given that the MUST APPEAR operator is usually thought of as an operator to decrease the number of retrieved results. The AND operator resulted in an increase in coverage on MSN but no change in coverage on AOL or Google. Again, this increase in coverage on MSN is surprising because the AND operator is typically used to decrease the number of retrieved results. Finally, the OR operator also had mixed results, with an increase in coverage on Google, but no change in coverage on AOL or MSN.

Returning to our hypotheses for research question one, we certainly cannot reject Hypothesis 1d. The PHRASE operator resulted in a decrease in coverage across all three search engines. Hypotheses 1a and 1c are rejected. The AND and the MUST APPEAR operator did not result in a decrease in coverage on any search engines. Hypothesis 1b is partially supported. The OR operator resulted in an increase in coverage on Google, but there was no change in coverage on AOL or MSN with the OR operator.

6.1.1 *Coverage by Search Engine.* We conducted a Scheffe's comparison of the coverage of the three search engines to highlight any significant differences among them. This comparison is presented in Table II.

Measuring the coverage effect of query operators on particular search engines is a subtle task, given the complex interplay of query terms, query operators, retrieval algorithms, and content collections. Using this set of queries, there was a significant difference between MSN and the other two search engines, but there was no significant difference between Google and AOL, as shown in Table II. MSN retrieved far more documents (total of 459,868,880; mean of 2,322,570) than the other two search engines. MSN retrieved almost 17 times as many documents as AOL (total of 27,555,106; mean of 147,353) and 6 times as many documents as Google (74,806,215; mean of 337,809).

6.2 Relative Precision

Of the 600 queries submitted, 570 retrieved 10 or more results. There were 13 queries that retrieved no results, and 17 queries that returned at least one but fewer than 10 results. In total, there were 5,748 documents retrieved by all the queries on all three search engines that were ranked in the top ten results for each query. There were 3,328 relevant documents and 2,420 non-relevant documents retrieved as evaluated by the four raters. Table III shows the number of documents retrieved by the original queries, the modified queries, and the combined set of queries.

Table III. Number of Results Retrieved by Query

Number of Results (Max. of 10)	Original Queries (with Operators)	Modified Queries (No Operators)	All Queries
10	287	283	570
9	0	0	0
8	0	0	0
7	0	1	1
6	1	1	2
5	1	0	1
4	0	0	0
3	1	1	2
2	2	5	7
1	1	3	4
0	7	6	13
Total Number of Queries	300	300	600

Table IV. Paired Sample t-test with and without Operators for Relative Precision across Three Search Engines

Operator	AOL	Google	MSN
AND	.88 (df = 24)	0.20 (df = 24)	−1.52 (df = 24)
Mean Without	.44	.78	.56
Mean With	.46	.79	.52
OR	−.36 (df = 24)	**−2.1* (df = 23)**	.82 (df = 23)
Mean Without	.494	.69	.467
Mean With	.486	.57	.500
MUST APPEAR	−.75 (df = 23)	1.37 (df = 24)	−.99 (df = 22)
Mean Without	.478	.66	.47
Mean With	.461	.72	.41
PHRASE	**2.85** (df = 24)**	−1.16 (df = 23)	1.96 (df = 23)***
Mean Without	.50	.81	.49
Mean With	.66	.74	.64

Notes: (1) $^{**}p < 0.01$, (2) $^{*}p < 0.05$, ***nearly significant at $p < 0.06$.

Our second research question was whether the use of query operators would result in an increase in relative precision. Our hypothesis was that the use of query operators would result in an increase in relative precision or P@10 (i.e., number of relevant documents in the top ten). The results of the paired-sample t-tests are shown in Table IV.

From Table IV, we see that the AND operator had no effect on relative precision across all three search engines. The OR operator had no effect on AOL, decreased relative precision on Google and had no effect on MSN. The MUST APPEAR operator had no effect on relative precision across all three search engines. The PHRASE operator increased the relative precision on AOL, but it had no effect on Google or MSN.

For our hypotheses for research question number 2, we reject hypotheses 2a and 2c: the AND and MUST APPEAR operators had no effect on relative precision on any of the three search engines. We also reject hypothesis 2b: the OR operator either had no effect or caused a decrease in relative precision. There was partial support for hypothesis 2d. The PHRASE operator resulted

Table V. Results of Pairwise Comparisons of Relative Precision by Search Engine

Search Engine Comparison	Difference Between Relative Precision Means	Simultaneous 95% Confidence Limits
Google—MSN	**0.20***	0.12–0.30
Google—AOL	**0.22***	0.13–0.31
MSN—AOL	0.02	−0.07–0.10

Note: (1) *$p < 0.05$.

Table VI. Paired Sample t-test with and without Operators for Ranking on AOL Search Engine

Rank	AND	OR	MUST APPEAR	PHRASE
1	−0.75 (df = 24)	−1.12 (df = 24)	0.33 (df = 24)	1.68 (df = 22)
2	**−2.41* (df = 24)**	−1.01 (df = 24)	−0.61 (df = 23)	1.42 (df = 22)
3	0.63 (df = 24)	1.80 (df = 21)	−1.42 (df = 23)	1.49 (df = 22)
4	1.32 (df = 24)	0.23 (df = 21)	1.87 (df = 23)	0.85 (df = 22)
5	−1.44 (df = 24)	0.32 (df = 21)	−0.38 (df = 23)	0.85 (df = 22)
6	0.96 (df = 24)	0.20 (df = 21)	−0.72 (df = 23)	0.41 (df = 22)
7	0.86 (df = 24)	−0.98 (df = 21)	0.25 (df = 23)	1.82 (df = 22)
8	1.98 (df = 24)	0.66 (df = 21)	0.10 (df = 23)	1.54 (df = 22)
9	1.07 (df = 24)	1.26 (df = 21)	−1.46 (df = 23)	1.58 (df = 22)
10	0.08 (df = 24)	−0.20 (df = 21)	−0.86 (df = 22)	0.96 (df = 22)

Note: (1) *$p < 0.05$.

in an increase in relative precision on AOL but had no effect on Google or MSN.

6.2.1 *Relative Precision by Search Engine.* We conducted a comparison of the three search engines to determine if there were any significant differences among them in terms of relative precision. There was a significant difference between Google and the other two search engines. There was no significant difference between MSN and AOL, as shown in Table V.

The average P@10 of Google was 0.73 using no operators and 0.68 with operators. For MSN, average P@10 was 0.49 without operators and 0.52 with operators. AOL's average P@10 was 0.47 using no operators and 0.52 with operators. Using Google, a searcher would normally retrieve approximately two additional relevant documents in the top ten, relative to the number of relevant documents retrieved using AOL or MSN. Given that the reviewers were aware of the underlying search engines retrieving the results, there is the possibility of bias in this analysis. However, the primary aim of this study is to evaluate the effect of query operators, not to compare and contrast search engine performance.

6.3 Ranking

Our third research question concerned the effect that query operators would have on the ranking of relevant documents. The results of our analysis are presented in Tables VI, VII, and VIII.

The only significant change in rankings was for the second position for AOL using the AND operator, and this was a decrease in precision at that rank.

Table VII. Paired Sample t-test with and without Operators for Ranking on Google Search Engine

Rank	AND	OR	MUST APPEAR	PHRASE
1	1.81 (df = 23)	−1.48 (df = 23)	**2.88**** **(df = 24)**	−0.52 (df = 23)
2	1.82 (df = 22)	−1.27 (df = 22)	1.56 (df = 24)	−1.36 (df = 23)
3	−1.09 (df = 22)	**−2.41*** **(df = 22)**	0.59 (df = 24)	−1.39 (df = 22)
4	−0.22 (df = 21)	**−3.00**** **(df = 22)**	−0.57 (df = 24)	−0.23 (df = 22)
5	−0.10 (df = 21)	**−2.65**** **(df = 22)**	0.21 (df = 24)	−1.49 (df = 22)
6	−0.07 (df = 21)	−1.67 (df = 22)	0.00 (df = 24)	−1.82 (df = 22)
7	0.47 (df = 21)	**−3.52**** **(df = 22)**	0.33 (df = 24)	−0.08 (df = 22)
8	0.76 (df = 21)	−0.76 (df = 22)	0.74 (df = 24)	−1.09 (df = 22)
9	−0.16 (df = 21)	−1.15 (df = 22)	0.19 (df = 24)	−0.13 (df = 22)
10	−0.21 (df = 21)	−1.61 (df = 22)	1.18 (df = 24)	−0.78 (df = 22)

Notes: (1) $^{**}p < 0.01$, $^{*}p < 0.05$.

Table VIII. Paired Sample t-test with and without Operators for Ranking on MSN Search Engine

Rank	AND	OR	MUST APPEAR	PHRASE
1	0.88 (df = 24)	**2.49*** **(df = 23)**	0.01 (df = 23)	0.75 (df = 23)
2	1.26 (df = 24)	0.08 (df = 22)	−0.59 (df = 22)	0.40 (df = 23)
3	−0.71 (df = 24)	0.61 (df = 22)	−0.97 (df = 22)	1.84 (df = 22)
4	1.26 (df = 24)	1.92 (df = 22)	−1.20 (df = 21)	**2.61*** **(df = 22)**
5	−1.94 (df = 24)	1.10 (df = 22)	−0.32 (df = 21)	**2.09*** **(df = 22)**
6	1.04 (df = 24)	−0.71 (df = 22)	−1.48 (df = 21)	0.44 (df = 22)
7	−0.39 (df = 24)	0.22 (df = 22)	−0.81 (df = 21)	**2.10*** **(df = 22)**
8	0.00 (df = 24)	1.47 (df = 22)	−1.51 (df = 21)	1.53 (df = 22)
9	0.00 (df = 24)	0.56 (df = 22)	0.25 (df = 21)	1.23 (df = 22)
10	0.50 (df = 24)	1.28 (df = 22)	−0.63 (df = 21)	0.93 (df = 22)

Note: (1) $^{*}p < 0.05$.

However, with forty t-tests, there is a random chance that one or two will be significant at the 0.05 level.

From Table VII, the MUST APPEAR operator resulted in an increase at rank 1 and the OR operator resulted in a decrease of relevant documents at ranks 3, 4, 5 and 7 using Google.

From Table VIII, the OR operator resulted in an increase at rank 1 and the PHRASE operator resulted in increases at ranks 4, 5 and 7 using MSN. Means for all three search engines at all ten rankings appear in Appendix C.

For research question number 3, we can reject hypothesis 3a. The AND operator resulted a decrease at rank 2 for AOL and had no effect on Google or MSN. The OR operator had no effect on AOL, resulted in a decrease in relevant documents at ranks 3, 4, 5, and 7 on Google, and an increase at rank 1 on MSN. The MUST APPEAR operator resulted in an increase in relevant documents at rank 1 for Google and had no effect on AOL or MSN. The PHRASE operator had no effect on AOL or Google rankings but resulted in an increase in ranks 4, 5, and 7 on MSN. Therefore, there is partial support for hypotheses 3b, 3c, and 3d depending on the search engine.

User studies show that ranking has a positive impact on the perceived performance of IR systems [Witten et al. 1994]. Given the importance of

Table IX. Number and Percentage of Relevant Results by Ranked Position and Query Type

Rank Position	Original Queries (with Operators)		Modified Queries (No Operators)		All Queries	
1	188	11%	183	11%	371	11%
2	185	11%	191	12%	376	11%
3	179	11%	179	11%	358	11%
4	182	11%	172	10%	354	11%
5	164	10%	163	10%	327	10%
6	160	10%	166	10%	326	10%
7	156	9%	153	9%	309	9%
8	165	10%	154	9%	319	10%
9	155	9%	151	9%	306	9%
10	142	8%	140	8%	282	8%
Total Results	1676	100%	1652	100%	3328	100%

Table X. Number and Percentage of Relevant Results by Ranked Position by Search Engine

Rank Position	AOL		Google		MSN	
1	97	10%	147	11%	127	13%
2	98	10%	153	11%	125	13%
3	108	11%	144	10%	106	11%
4	109	11%	142	10%	103	10%
5	89	9%	140	10%	98	10%
6	95	10%	135	10%	96	10%
7	92	10%	134	10%	83	8%
8	90	9%	139	10%	90	9%
9	87	9%	135	10%	84	8%
10	85	9%	115	8%	82	8%
Total	950	100%	1384	100%	994	100%

ranking, we discuss our findings concerning ranking further in the next three sections.

6.3.1 *Documents in Ranking.* As stated previously, the queries retrieved 5,748 documents. Results that appeared higher than position 10 in the results list were not utilized in the comparison of rankings. Table IX presents data on the number of relevant results returned and reported by ranked position.

Table IX shows that there was very little change in ranking at any position, regardless of whether or not the modified or original queries were utilized. There were 3,328 relevant results returned. The original queries returned 1,676 (50.36%) relevant results, and the modified queries returned 1,652 (49.64%) relevant results.

6.3.2 *Ranking by Search Engine.* We conducted a comparison of ranking for each search engine. These results are displayed in Table X.

As shown in Table X, there was very little variance in percentages among the three search engines at each rank. Google was higher in absolute numbers, reflecting Google's higher P@10 compared to the other two search engines. MSN

Table XI. Number and Percentage of Relevant Results by Ranked Position by Operator

Rank Position	AND		OR		MUST APPEAR		PHRASE	
1	99	11%	78	10%	94	12%	100	11%
2	105	12%	77	10%	93	12%	101	11%
3	100	11%	84	11%	83	11%	91	10%
4	94	11%	84	11%	81	10%	95	10%
5	93	11%	79	10%	63	8%	92	10%
6	85	10%	75	10%	78	10%	88	10%
7	81	9%	68	9%	72	9%	88	10%
8	77	9%	73	10%	76	10%	93	10%
9	82	9%	70	9%	73	9%	81	9%
10	69	8%	66	9%	64	8%	83	9%
Total	885	100%	754	100%	777	100%	912	100%

had a higher percentage of relevant documents ranked in the top two positions, relative to AOL and Google.

6.3.3 *Ranking by Query Operator.* We also conducted a comparison for each search operator. These results are displayed in Table XI.

As illustrated in Table XI, there was little change in percentage of relevant documents at each rank among the four query operators. The absolute numbers for the PHRASE operator are higher, reflecting its greater impact on relative precision. The MUST APPEAR operator had a slightly higher number of relevant documents ranked in positions one and two relative to the other three operators. Appendix D displays graphical versions of the data contained in Tables IX, X, and XI.

7. DISCUSSION OF RESULTS

We discuss our results first in the areas of coverage, relative precision, and ranking. We then discuss them from an overall perspective. Since we are examining the behavior of proprietary search engines whose algorithms and operations are only partially disclosed to the public, much of this discussion treats the search engines as black boxes (or at least dark gray boxes). However, these types of "gray box" studies (e.g., Ding and Marchionini [1996], Hawking et al. [2001], Lawrence and Giles [1999], and Spink et al. [2002]) are very beneficial in illuminating the complex interactions occurring on the Web in ways that are difficult to recreate in controlled studies [Dumais 2002].

Both search engine and query operator had a significant effect on coverage, which is the total number of documents found. Of course, these hits correspond to the matching documents identified by the system; they were not actually all retrieved and presented to the users. MSN retrieved significantly more documents than either AOL or Google. 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. However, there is no evidence that MSN is retrieving more documents because its database is significantly larger; it may simply be using a lower threshold for matching. It could also be that MSN uses a more liberal estimation algorithm relative to AOL or Google.

During the time of the study, AOL utilized Inktomi, and MSN used a combination of Web directories and Inktomi listing, depending on the particular query [Notess 2003; Sullivan 2002]. We reasoned that a switch from using solely the Web directory content to the Inktomi listing would usually result in dramatic increase in the number of results on MSN. We analyzed the coverage numbers in order to isolate these possible queries. For basic and advanced versions of the same query, the difference in coverage was usually within a range of 6% or less. However, there were nineteen queries above this threshold, two queries on AOL and seventeen on MSN. On MSN, the increase from the seventeen queries was the result of more retrieved Web pages (i.e., documents outside the MSN directory content).

The PHRASE operator was the only operator that results in a decrease in coverage across all three search engines. The AND and MUST APPEAR operators are also conventionally used to narrow a query, which would be expected to result in lower coverage, although they did not do so in our analysis. In fact, on MSN, the AND and MUST APPEAR operators resulted in an increase in coverage. The OR operator is more inclusive and generally used to broaden a query, and this was its effect on Google, which is consistent with its statement that AND is its default operator. The OR operator had no significant effect on the coverage of AOL or MSN. It was interesting to note that, despite the statement that AND is the default operator on Google, searches using AND and searches using NO OPERATOR did not always return identical results lists, even when repeated. Some queries did not conform to the stated or expected effect. As examples, on Google the following queries did not return identical results or identical numbers of results (1) *word* AND *search* versus *word search* and (2) *+pc +to +phone +calls* versus *pc to phone calls*. With the MUST APPEAR query, the MUST APPEAR operator forced the inclusion of the stop word *to*, which probably caused the change in retrieved results.

It is interesting that there is no significant decrease in coverage using the AND and MUST APPEAR operators. Many would assume that the use of these operators would narrow the query. Our results indicate that they are of no assistance in this regard. Conversely, it is interesting there is no increase in coverage using the OR operator on AOL or MSN, as one would expect this operator to broaden the query. This is probably the result of the use of some version of OR as the default operator in these search engines.

Both search engine and operator had a significant impact on relative precision, although with query operators the impact was not always an improvement. Google's performance in relative precision was statistically significantly higher than that of AOL or MSN (i.e., about two additional relevant documents out of ten documents). It is likely that this superior performance is a result of superior matching and ranking algorithms rather than database size since all three have substantial databases. Of course, relative performance rankings observed at a particular time may well change as search engines develop and implement improved algorithms for indexing, matching, and ranking or change providers of these services. Since AOL has recently switched from Inktomi to Google for backend support, its performance today may not be significantly different from that of Google.

Table XII. Comparison of Relative Precision by Result Pages for Original and Modified Queries

Search Engine	Number of Results Pages Where Modified Queries (i.e., No Operators) Retrieved			Total Results Lists
	Same Number of Relevant Results	Fewer Relevant Results	More Relevant Results	
AOL	40	36	24	100
Google	58	13	29	100
MSN	45	28	27	100
Total	143 (47.7%)	77 (25.7%)	80 (26.7%)	300

For the most part, specific query operator choice did not have a significant improvement on relative precision. However, the PHRASE operator did have significantly better performance on AOL and nearly significant on MSN. The OR operator resulted in a decrease in performance on Google. The interpretation and implementation of these operators (and others) probably has more impact on relative precision than the logical definitions. Logically, the OR operator would result in the broadest query, and the PHRASE operator the narrowest. However, if the search engine were to rank items with the terms appearing in a phrase higher than other items, one would expect the top listings to be similar, if not identical.

One of the conclusions from our results is that the precision anomalies identified by Eastman [2002] do in fact occur fairly frequently. Such anomalies result when query modifications—intended to improve query precision—in fact have no impact or actually *decrease* relative precision. The average precision across all queries and search engines showed that precision was higher for the simple queries about as often as for the advanced queries and that usually there was no change at all. Table XII illustrates the commonality of results listings encompassing the top ten results across search engines for all query operators.

From Table XII, queries with and without operators retrieved the same number of relevant results approximately 48% of the time. Perhaps even more surprisingly, approximately 27% of the time, queries without operators retrieved more relevant results than queries with operators. Neither search engine nor query operator had a significant impact on ranking within the top 10 items. These results augment previous research showing that there is a high degree of overlap in results when searching with query operators and without query operators [Jansen 2000]. Not only are the results similar, but also there is no improvement in the relevance of the new documents or in ranking of those relevant documents. Overall, Table XII illustrates that for approximately 74% of queries, searchers obtain no improvement from using query operators.

Another conclusion from this analysis is that a user cannot effectively use query operators without an understanding of the underlying IR system. Logically, we can define how we expect them to perform; but when implemented on real systems, the results are sometimes not intuitive, such as when operators generally thought of as narrowing queries actually broaden them on certain search engines. One wonders if it is realistic to expect this detailed system knowledge of searchers?

Of the three performance metrics considered here, we expect relative precision to have the most real impact and importance to users. Coverage provides an interesting perspective on performance, but increased coverage is unlikely to provide additional benefit to a user who is typically only going to examine the top ranked items in any case. Improvement in coverage is of limited value to most searchers. We acknowledge, of course, that within some domains, coverage will be of importance. From a research and system design perspective, coverage is of importance with implications for users, system designers, and content providers. Research on the relationship between subsets of a collection and a full collection has gone on for some time [Salton and McGill 1983] and continues today [Hawking and Robertson 2003]. The recent work by Hawking and Robertson [2003] indicate that increased collection size should have a positive impact on Web search engine performance relative to a sub-sample of that collection.

Choice of search engine and, to a limited extent, use of operators did have an impact on relative precision. It appears that there is little advantage to using OR in a query, but there may be an advantage, at least in some cases, in using the PHRASE operator. A difference in ranking might be expected to make some difference to the user since it is more convenient to have relevant items at the top of the list. However, this study found only spotty improvements to ranking with no general improvement using any operator.

Concerning the strengths of our research, the first is that the original queries we used represent real needs of real users that they submitted to a real Web search engine. These users employed query operators in the manner they did in order to improve Web search engine results. These queries are a sample of how users actually employ query operators. Second, we utilized three of the most popular search engines on the Web as measured by number of unique visitors. Our results have relevance for the substantial portion of Web searchers using these search engines. Finally, we evaluated a large number of queries and documents while controlling for query terms. This methodology permitted us to isolate the effects of the operators across a large range of topics.

As with any study, there are limitations. First, we did not investigate queries with occurrences of multiple operators, the Boolean NOT operator, or the MUST NOT APPEAR operator. Perhaps more sophisticated queries would result in more dramatic changes in search engine results. One would need to investigate whether or not these queries would produce any increase in relevance. Second, all our queries were originally submitted to the Excite search engine, introducing the possibility that these queries do not represent techniques of the general Web search engine user population. However, Jansen and Pooch [2001] have shown users of Web search engines exhibit common characteristics. Third, we used three search engines among the thousands available on the Web. Perhaps other search engines would have performed differently. Indications from previous research suggest, however, that many of these other general-purpose search engines (e.g., Alta Vista, Excite, Northern Light, AlltheWeb.com) would have performed similarly [Eastman 2002; Jansen 2000]. There has been little study of niche and content specific search engines. Fourth, the evaluators

were aware of the underlying systems when examining queries. It may have little effect when measuring the effect of query operators, but it could bias the comparisons among search engines. Finally, using real-world systems entails a lack of control relative to lab and other controlled studies. For example, AOL and MSN use multiple sources for documents, which may impact our research. Other hardware and software issues (i.e., cache storage, servers down, changing indexes [Sullivan 2003]) may impact the results. Concerning coverage measurements, all three search engines use heuristic algorithms to estimate the number of results. The algorithms may vary among the three. However, using real world systems captures the multivariable complexity of information seeking on the Web [Dumais 2002].

8. CONCLUSIONS AND FUTURE RESEARCH

Indexing and searching have been described as parallel and complementary activities, in which indexers attempt to determine which terms searchers will use, and searchers attempt to determine which terms the indexers have used. Web search engine design and Web searching constitute similar, but more complex, parallel activities. Search engine design involves not only the indexing algorithms, but also the matching and ranking algorithms. Web searching involves not only selection of terms, but also the construction of queries and selection of search engines. So, future work in this area involves research from the perspectives of both the search engine designer and the search engine user. Our focus here has been primarily from the perspective of the user, namely: If certain operators are used, will the results be any better?

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. However, this study could be easily expanded to address some issues related to coverage by including more queries since evaluation of coverage, unlike evaluation of relevance and ranking, does not require user judgments. In addition to a larger sample of queries similar to those used here, queries using additional operators or no operators could be included. Variations in coverage resulting from use of different interface and search options could also be considered, along with whether or not increases in coverage result in greater relative precision.

The results of this research contribute to our understanding of how these real world Web IR systems are employed, highlight the general state of the art when combined with other prior research, and provide a rationale for exploration of other research areas to improve the design of Web IR systems. The help pages of many Web search engines would lead one to believe that the use of query operators would increase searching effectiveness. The results of this study indicate that generally they provide little or no benefit. Moreover, they are counter productive in some cases.

The implications for education and training in information retrieval are fairly clear, especially in view of the growing body of research on user behavior. Most users are doing fine with Web searches without using complex search

strategies and operators. The results of this study indicate that this behavior is reasonable since using such operators does not in general significantly improve the results of searches. The amount of training and practice that would be needed to enable most users to correctly formulate and use advanced operators, along with the apparent need to understand the particular IR system, is not justified by the relatively small potential improvement in results. So, training and experience in more sophisticated searching techniques and strategies could reasonably be limited to information professionals who might be expected to have a use for them, are knowledgeable on a particular system or set of systems, and engage in intricate searching tasks.

While it appears to be the case that most users are satisfied with at least some searches and their results [Spink et al. 1998a], it is not true that all searchers are satisfied with all search results. We need more focused information on the topics and situations that might lead to unsatisfactory results. It would also be useful to know how often users might be satisfied with results which are, by some objective standard, simply incorrect or misleading.

The recommendation that most users of Web search engines should not be expected to use advanced operators does not necessarily imply that such operators no longer have a place in searching. Only general search engines targeted at the general public were considered here. There are IR systems that do not have very sophisticated matching and ranking algorithms for which advanced operators may be needed to achieve satisfactory results. Also, some types of searches, such as many legal searches, have very high recall requirements, unlike the searches considered here.

There are system design and research implications of this study. 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 are several promising research areas already underway, such as content [Deerwester et al. 1990] and link analysis [Brin 1998]. Other researchers are exploring the use of metadata [Craswell et al. 2001] and term feedback [Hiemstra and Robertson 2001]. Personalization [Jansen and Kroner 2003] and niche search engines [Glover et al. 2001] also have the potential to improve IR performance. Though personalization, for example, the system could perhaps automatically detect when the use of Boolean operators would be effective. The system could implement the operators for the user, aiming for a more effective query [Cronen-Townsend et al. 2002]. This is a finer granularity of what Google does by defaulting to AND in order to improve precision.

Another aspect of the search process that might be considered is the user interface. The user interfaces of current search engines often require or encourage mode switching between basic and advanced modes, which is generally regarded as undesirable. The advanced searching screens generally do not directly support a full range of Boolean queries, including queries using both AND and OR. So, these interfaces appear to be more complex than most users need and both more complicated and more limited than sophisticated and experienced searchers might want. Research into interface design is of obvious importance for both entering the query and displaying the results [Dumais et al. 2001].

These comments about the desirability of advanced query operators for most Web users also do not imply that consideration and use of term relationships is not important. However, these are considerations that can be handled by the system rather than the user. We can now build IR systems that, unlike the first computerized IR systems, implement ranking algorithms that take the place of sophisticated user queries. Although such ranking systems have been investigated and built for many years, it has only been with the advent of Web search engines that they have come into widespread use. Given a short list of query terms, it is easy to implement a system which ANDs them together or converts them to a phrase and then ranks items higher if they match the narrower interpretations. Not only is it straightforward to implement algorithms which do the work that would be needed to construct an appropriate Boolean query, but far more sophisticated algorithms have been developed using linkage and other information not available to the user for query construction. The balance of work between the user and the IR system is shifting from that which existed for the first computerized IR systems; more of the work is being performed by the system. So, the design and implementation of appropriate algorithms for matching and ranking continues to be of critical importance.

APPENDIXES

A. Queries Utilized for Research

Table XIII. List of Queries Used in Research

Original Queries Containing Query Operators Submitted by Search Engine Users	Simplified Query (i.e., Query Operators Removed) Used for Comparison
daktarin AND nail	daktarin nail
wallpaper AND nature scenes	wallpaper nature scenes
shoes AND sandals	shoes sandals
pocket pc AND free games	pocket pc free games
word AND search	word search
french revolution AND jews	french revolution jews
word games AND boggles	word games boggles
baby furniture AND charlotte	baby furniture charlotte
nbc AND news	nbc news
cats AND kittens	cats kittens
job description AND salary on the study of gynecology	job description salary on the study of gynecology
stress AND humor	stress humor
hevia AND bagpipe	hevia bagpipe
car AND driver AND reports	car driver reports
saturn AND 2001 AND sl AND ratings	saturn 2001 sl ratings
saturn AND s-series AND ratings	saturn s-series ratings
java AND xml	java xml
aussie AND scans	aussie scans
architectures AND geometry	architectures geometry
baker AND don	baker don
hot wheels AND storage	hot wheels storage

Continued

Table XIII. Continued

Original Queries Containing Query Operators Submitted by Search Engine Users	Simplified Query (i.e., Query Operators Removed) Used for Comparison
minorities AND Alzheimer	minorities Alzheimer
michael lipman AND artist	michael lipman artist
silver AND hydrogen AND electrode AND diagram	silver hydrogen electrode diagram
china AND plane AND images	china plane images
java OR XML	java XML
ann OR kristin OR bondhus	ann kristin bondhus
computer OR internet providers	computer internet providers
school OR education OR university OR primary	school education university primary
animals OR plants	animals plants
fleetwood OR 25w	fleetwood 25w
imail OR 6 OR crack	imail 6 crack
defamation OR canadian OR law	defamation canadian law
AutoCAD OR training	AutoCAD training
driver OR for OR audio OR excel OR cmi OR 8838	driver for audio excel cmi 8838
chisinau OR moldovia	chisinau moldovia
microsoft OR office OR 2000	microsoft office 2000
game OR boy OR advance OR rom	game boy advance rom
www OR burodromes	www burodromes
mob OR mafia movies	mob mafia movies
sopranos wallpaper OR desktop themes	sopranos wallpaper desktop themes
pausen OR testament	pausen testament
hockey OR savannah	hockey savannah
password OR search	password search
aim OR estone	aim estone
radio OR receiver OR design	radio receiver design
heath OR electronics	heath electronics
vacation OR spots OR in OR southern OR texas	vacation spots in southern texas
free OR banner OR download	free banner download
vidsdiv3 OR codec OR download	vidsdiv3 codec download
“insightful quotes”	insightful quotes
“mary kay letourneau”	mary kay letourneau
“form 1003”	form 1003
“bargain music”	bargain music
“camp jeep”	camp jeep
“zurich capital”	zurich capital
“travis street”	travis street
“travis street partners”	travis street partners
“mp3 abstract”	mp3 abstract
“trac phone”	trac phone
“university of texas” ”alternative medicine”	university of texas alternative medicine
“gale research company”	gale research company
“norwich football”	norwich football
“nambu type 94”	nambu type 94
“battery belt”	battery belt
“favorite away messages”	favorite away messages
“johnny got his gun”	johnny got his gun
“gl doom”	gl doom

“chevy kodiak”	chevy kodiak
“music equipment”	music equipment
“fender guitars”	fender guitars
“music equipment”	music equipment
“custom tassels”	custom tassels
“ultrasonic repeller”	ultrasonic repeller
“tax law and depreciation system”	tax law and depreciation system
+las +vegas	las vegas
+pennsylvania +real +estate	pennsylvania real estate
+heterogenous +database +problem	heterogenous database problem
+ars +tubes	ars tubes
+gemstone +madeira +citrene +refraction	gemstone madeira citrene refraction
+madeira citrene +gemstone	madeira citrene gemstone
+music +skeets	music skeets
+oceania +art	oceania art
+pc +to +phone +calls	pc to phone calls
+crane +greer	crane greer
+coolermaster +review +case	coolermaster review case
+furniture +moving +equipment	furniture moving equipment
+homeschooling +tutoring	homeschooling tutoring
+schwabe +transformer	schwabe transformer
+arab +countries	arab countries
+alaska +airlines +web +specials	alaska airlines web specials
+bose +acoustimass	bose acoustimass
+bose +lifestyle +used equipment	bose lifestyle used equipment
+aquarium +salt +ny	aquarium salt ny
+lyrics +offenbach +barcarolle	lyrics offenbach barcarolle
+lou +gehrig +disease	lou gehrig disease
+male +pattern +baldness	male pattern baldness
+l3Enc +download	l3Enc download
+scooter +invention +ginger	scooter invention ginger
+borland +software +founder	borland software founder

B. Instructions to Reviewers

Do the following for each of the queries/searches:

Determine which retrieved items (top 10 only, in case there are more) are relevant to the query. If you are not sure, use your best judgment. (However, feel free to comment if you have no idea what the topic is about.) Indicate the degree of relevance using the following scale given below. Write the degree of relevance on the printout beside the item. Also, include your evaluator number (to be assigned later) at the top of each page.

4 – Totally relevant
3 – Mostly relevant
2 – Marginally relevant
1 – Not relevant

C. Mean Number of Relevant Documents at Each Rank

Table XIV. Mean Number of Relevant Documents with and without Operators for Ranking on AOL

Rank	AND Mean With / Mean Without	OR Mean With / Mean Without	MUST APPEAR Mean With / Mean Without	PHRASE Mean With / Mean Without
1	2.72	2.49	2.83	3.20
	2.82	2.72	2.78	2.74
2	**2.85**	2.49	2.80	3.01
	3.09*	2.71	2.89	2.78
3	2.90	2.98	2.51	3.18
	2.83	2.69	2.71	2.91
4	3.06	2.70	2.77	3.19
	2.91	2.69	2.57	3.01
5	2.65	2.68	2.47	3.13
	2.74	2.63	2.51	2.97
6	2.82	2.51	2.55	2.97
	2.73	2.48	2.61	2.88
7	2.63	2.46	2.68	3.00
	2.56	2.55	2.65	2.63
8	2.61	2.55	2.63	3.08
	2.43	2.46	2.62	2.80
9	2.74	2.64	2.43	2.95
	2.62	2.48	2.62	2.63
10	2.52	2.49	2.57	2.99
	2.51	2.51	2.68	2.78

Note: (1) $^{*}p < 0.05$.

Table XV. Mean Number of Relevant Documents with and without Operators for Ranking on Google

Rank	AND Mean With / Mean Without	OR Mean With / Mean Without	MUST APPEAR Mean With / Mean Without	PHRASE Mean With / Mean Without
1	3.62	2.80	**3.35**	3.18
	3.40	3.05	**3.07****	3.27
2	3.60	2.71	3.35	3.17
	3.43	2.93	3.27	3.33
3	3.33	**2.84**	3.20	3.07
	3.46	**3.24**$^{*\#}$	3.18	3.26
4	3.45	**2.90**	3.10	3.20
	3.48	**3.27**$^{**\#}$	3.14	3.24
5	3.39	**2.92**	2.89	3.16
	3.40	**3.25**$^{**\#}$	2.86	3.34
6	3.25	2.84	3.07	3.00
	3.26	3.10	3.07	3.27
7	3.36	**2.73**	3.17	3.24
	3.31	**3.20**$^{**\#}$	3.15	3.25
8	3.25	2.99	3.22	3.11
	3.17	3.10	3.16	3.25
9	3.26	2.84	2.98	3.13
	3.28	3.06	2.96	3.15
10	2.99	2.59	2.88	3.20
	3.02	2.95	2.73	3.28

Notes: (1) $^{*}p < 0.05$, (2) $^{**}p < 0.01$, (3) $^{\#}$Decrease in mean number of relevant documents.

Table XVI. Mean Number of Relevant Documents with and without Operators for Ranking on MSN

Rank	AND Mean With Mean Without	OR Mean With Mean Without	MUST APPEAR Mean With Mean Without	PHRASE Mean With Mean Without
1	3.26 3.15	**3.00*** **2.80**	2.88 2.54	3.20 3.05
2	3.19 3.02	2.84 2.84	2.78 2.86	3.24 3.15
3	3.00 3.15	2.73 2.65	2.59 2.74	3.10 2.65
4	2.92 2.75	3.02 3.82	2.61 2.78	**3.15*** **2.52**
5	2.89 3.03	2.66 2.53	2.60 2.65	**3.05*** **2.59**
6	2.68 2.59	2.52 2.63	2.73 2.93	2.66 2.55
7	2.67 2.71	2.53 2.50	2.30 2.43	**3.00*** **2.52**
8	2.66 2.66	2.58 2.33	2.49 2.72	2.92 2.57
9	2.59 2.59	2.52 2.46	2.63 2.60	2.73 2.44
10	2.80 2.76	2.64 2.40	2.40 2.50	2.67 2.44

Note: (1) * $p < 0.05$.

D. Number of Relevant Results by Rank

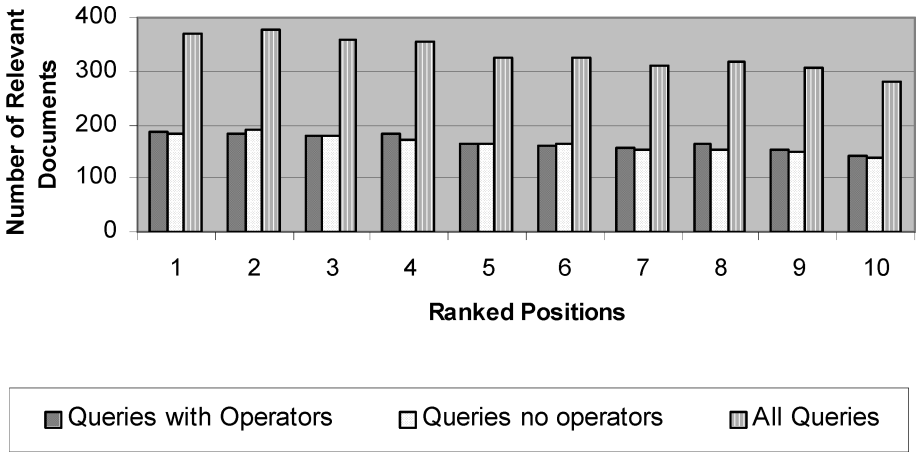

Fig. 1. Number of relevant results by rank and query type.

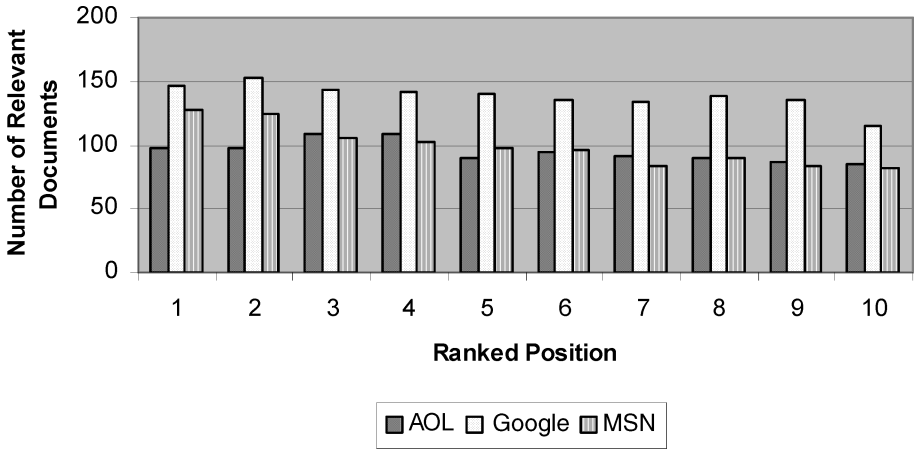

Fig. 2. Number of relevant results by rank and search engine.

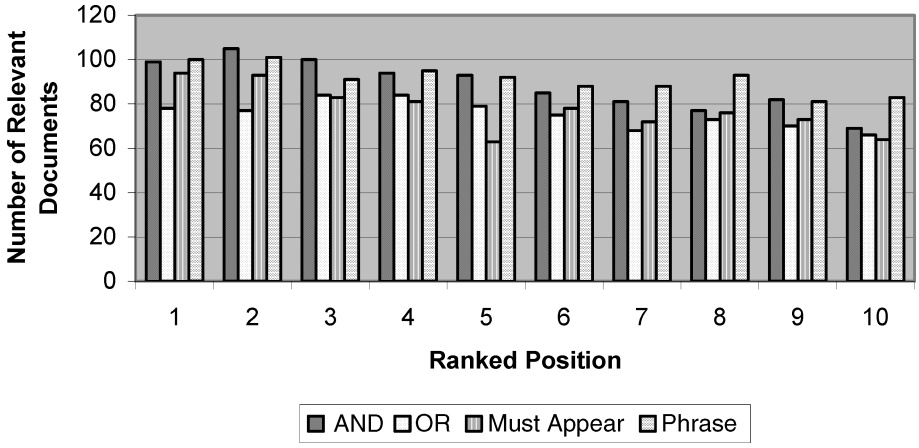

Fig. 3. Number of relevant results by rank and operator.

ACKNOWLEDGMENTS

We would like to thank Gail Amaker and Georgia Gibson for their contributions to this research project under the sponsorship of the South Carolina Alliance for Minority Participation. We also thank Doug Cutting, Jack Xu, and Soo Young Rieh from Excite.

REFERENCES

AOL. 2003. Getting started. Accessed on: 23 April 2003 at: http://search.aol.com/aolcom/help.jsp.

BORGMAN, C. 1996. Why are online catalogs still hard to use? *J. ASIS 47*, 7, 493–503.

BRIN, S. 1998. Extracting patterns and relations from the world wide web. In *Proceedings of the World Wide Web and Databases* (Valencia, Spain). 172–183.

CHANG, C.-C. K., GARCIA-MOLINA, H., AND PAEPCKE, A. 1999. Predicate rewriting for translating Boolean queries in a heterogeneous information system. *ACM Trans. Inf. Syst. 17*, 1, 1–39.

CHOWDHURY, A., BEITZEL, S., AND JENSEN, E. 2002. Analysis of combining multiple query representations with varying lengths in a single engine. In *Proceedings of the IEEE 3rd International Conference on Information Technology Coding and Computing* (Las Vegas, Nev.). IEEE Computer Society Press, Los Alamitos, Calif., 8–15.

CLARK, P. 2001. Solving Internet overload. *Net Econ. 2*, 3, 1.

CLARKE, C. L. A. AND CORMACK, G. V. 2000. Shortest-substring retrieval and ranking. *ACM Trans. Inf. Syst. 18*, 1, 44–78.

COOPER, W. S. 1968. Expected search length: A single measure of retrieval effectiveness based on weak ordering action of retrieval systems. *J. ASIS 19*, 1, 30–41.

CRASWELL, N., HAWKING, D., AND ROBERTSON, S. 2001. Effective site finding using link anchor information. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (New Orleans, La). ACM New York, 250–257.

CRONEN-TOWNSEND, S., ZHOU, Y., AND CROFT, W. B. 2002. Predicting query performance. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Tampere, Finland). ACM New York, 299–306.

CYBER ATLAS. 2002. U.S. Top 50 Internet properties July 2002 at home and work combined. Accessed on: 1 November 2002 at: http://cyberatlas.internet.com.

DEERWESTER, S., DUMAIS, S. T., FURNAS, G. W., LANDAUER, T. K., AND HARSHMAN, R. 1990. Indexing by latent semantic analysis. *J. ASIS 41*, 6, 391–407.

DING, W. AND MARCHIONINI, G. 1996. A comparative study of Web search service performance. In *Proceedings of the 59th Annual Meeting of the American Society for Information Science* (Medford, N.J.). 136–142.

DUMAIS, S., CUTRELL, E., and CHEN, H. 2001. Optimizing search by showing results in context. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Wash.). 277–284.

DUMAIS, S. T. 2002. Web experiments and test collections. Presented at the 11th International World Wide Web Conference (Honolulu, Hi., May 7–11). Presentation available at http://www2002.org/presentations/dumais.pdf.

EASTMAN, C. M. 2002. 30,000 hits may be better than 300: Precision anomalies in Internet searches. *J. ASIST 53*, 11, 879–882.

FORD, N., MILLER, D., AND MOSS, N. 2003. Web search strategies and approaches to studying. *J. ASIST 54*, 6, 473–489.

FRANTS, V. I., SHAPIRO, J., TAKSA, I., AND VOISKUNSKII, V. G. 1999. Boolean search: Current state and perspectives. *J. ASIS 50*, 1, 86–95.

GLOVER, E. J., LAWRENCE, S., GORDON, M. D., BIRMINGHAM, W. P., AND GILES, C. L. 2001. Web search—Your way: Improving Web searching with user preferences. *Commun. ACM 44*, 12, 97–102.

GOOGLE. 2003. Advanced search. Accessed on: 23 April 2003 at: http://www.google.com/help/refinesearch.html.

GUDIVADA, V. N., RAGHAVAN, V. V., GROSKY, W. I., AND KASANAGOTTU, R. 1997. Information retrieval on the World Wide Web. *IEEE Internet Comput. Sept.–Oct.* 58–68.

HAWKING, D. 2000. Overview of the TREC-9 Web track. In *Proceedings of the 9th Text Retrieval Conference* (Gaithersburg, Md.). 87–103.

HAWKING, D., CRASWELL, N., BAILEY, P., AND GRIFFIHS, K. 2001. Measuring search engine quality. *Inf. Ret. 4*, 1, 33–59.

HAWKING, D. AND ROBERTSON, S. 2003. On collection size and retrieval effectiveness. *Inf. Ret. 6*, 1, 99–105.

HIEMSTRA, D. AND ROBERTSON, S. E. 2001. Relevance feedback for best match term weighting algorithms in information retrieval. In *Proceedings of Joint 2nd DELOS-NSF Workshop on Personalisation and Recommender Systems in Digital Libraries* (Dublin City University, Ireland). 37–42.

HIEMSTRA, D. AND VRIES, A. D. 2000. CTIT Technical Report TR-CTIT-00-09: Relating the new language models of information retrieval to the traditional retrieval models. Accessed on: 30 March 2003 at: http://wwwhome.cs.utwente.nl/~hiemstra/papers/.

HÖLSCHER, C. AND STRUBE, G. 2000. Web search behavior of Internet experts and newbies. *Int. J. Comput. Telecommun. Netw. 33*, 1–6, 337–346.

INKTOMI. 2002. Refocused Inktomi seeks to monetize search. Accessed on: 20 December 2002 at: http://www.inktomi.com/company/news/.

JANSEN, B. J. 2000. An investigation into the use of simple queries on Web IR systems. *Inf. Res.: An Elect. J. 6*, 1, 1–10.

JANSEN, B. J. AND KRONER, G. 2003. The impact of automated assistance on the information retrieval process. In *Proceedings of the ACM CHI 2003 Conference on Human Factors in Computing Systems* (Fort Lauderdale, Fla.). ACM New York, 1004–1006.

JANSEN, B. J. AND POOCH, U. 2001. Web user studies: A review and framework for future work. *J. ASIST 52*, 3, 235–246.

JANSEN, B. J., SPINK, A., BATEMAN, J., AND SARACEVIC, T. 1998. Real life information retrieval: A study of user queries on the Web. *SIGIR Forum. 32*, 1, 5–17.

JANSEN, B. J., SPINK, A., PFAFF, A., AND GOODRUM, A. 2000a. Web query structure: Implications for IR system design. In *Proceedings of the 4th World Multiconference on Systemics, Cybernetics and Informatics (SCI'2000)*, (Orlando, Fla.). 169–176.

JANSEN, B. J., SPINK, A., AND SARACEVIC, T. 2000b. Real life, real users, and real needs: A study and analysis of user queries on the Web. *Inf. Proc. Manage. 36*, 2, 207–227.

KIRSCH, S. 1998. The future of Internet search (keynote address). Accessed on: 16 August 1999 at: http://www.skirsch.com/stk.html/presentations/sigir.ppt.

KORFHAGE, R. 1997. *Information Storage and Retrieval.* Wiley, New York.

LAWRENCE, S. AND GILES, C. L. 1999. Accessibility of information on the Web. *Nature. 400*, 107–109.

LEXISNEXIS. 2003. Terms and connectors—introduction to traditional Boolean searching. Accessed on: 5 April 2003 at: http://support.lexis-nexis.com/online/record.asp?ArticleID=GS_Boolean.

LOOKSMART. 2003. Looksmart search: About us. Accessed on: 1 February 2002 at: http://aboutus.looksmart.com/about.jhtml;$sessionid$MFR0PZYAAACADLAQQBTSOJQ?dir = profile&page = about.

LUCAS, W. AND TOPI, H. 2002. Form and function: The impact of query term and operator usage on Web search results. *J. ASIST 53*, 2, 95–108.

MSN. 2003. Tips for an advanced search. Accessed on: 23 April 2003 at: http://search.msn.com.

NIELSEN NETRATING. 2002. Top Web properties march 2002. Accessed on: 27 September 2002 at: http://www.nielsen-netratings.com.

NIELSENMEDIA. 1997. Search engines most popular method of surfing the Web. Accessed on: 30 August 2000 at: http://www.commerce.net/news/press/0416.html.

NOTESS, G. R. 2003. Search engine features chart. Accessed on: 13 April 2003 at: http://www.searchengineshowdown.com/features/.

PETERSEN, R. E. 1997. Eight Internet search engines compared. *First Monday. 2*, 2,

SALTON, G. AND MCGILL, M. J. 1983. *Introduction to Modern Information Retrieval.* McGraw-Hill, New York.

SELBERG, E. AND ETZIONI, O. 2000. On the instability of Web search services. In *Proceedings of RIAO 2000: Computer-Assisted Information Retrieval* (Paris, France). 223–236.

SILVERSTEIN, C., HENZINGER, M., MARAIS, H., AND MORICZ, M. 1999. Analysis of a very large Web search engine query log. *SIGIR Forum. 33*, 1, 6–12.

SORMUNEN, E. 2000. A novel method for the evaluation of Boolean query effectiveness across a wide operational range. In *Proceedings of 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Athens, Greece). ACM New York, 25–32.

SPINK, A. 1995. Term relevance feedback and mediated database searching: Implications for information retrieval practice and design. *Inf. Proc. Manage. 31*, 6, 161–171.

SPINK, A., BATEMAN, J. AND JANSEN, B. J. 1998a. Searching heterogeneous collections on the Web: Behavior of Excite users. *Inf. Res. 4*, 2, 317–328.

SPINK, A., BATEMAN, J. AND JANSEN, B. J. 1999. Searching the Web: A survey of Excite users. *J. Internet Res. Electron. Netw. Appl. Policy 9*, 2, 117–128.

SPINK, A., GREISDORF, H. AND BATEMAN, J. 1998b. From highly relevant to not relevant: Examining different regions of relevance. *J. ASIS 34*, 5, 599–621.

SPINK, A., JANSEN, B. J., WOLFRAM, D., AND SARACEVIC, T. 2002. From e-sex to e-commerce: Web search changes. *IEEE Comput. 35*, 3, 107–111.

SULLIVAN, D. 2000. Search watch. Accessed on: 1 June 2000 at: http://searchenginewatch.com/.
SULLIVAN, D. 2002. Search engine math. Accessed on: 11 April 2003 at: http://www.searchenginewatch.com/facts/article.php/2156021.
SULLIVAN, D. 2003. Coping with GDS, the Google dance syndrome. Accessed on: 14 July 2033 at: http://www.searchenginewatch.com/sereport/article.php/2216081.
WITTEN, I., MOFFAT, A. AND BELL, T. C. 1994. *Managing Gigabytes: Compressing and Indexing Documents and Images*. Van Nostrand Reinhold, New York, NY.
WOLFRAM, D. 1999. Term co-occurrence in Internet search engine queries: An analysis of the Excite data set. *Canad. J. Inf. Library Sci. 24*, 2/3, 12–33.
ZAPUR, K. AND ZHANG, J. 2000. Searching the Web using synonyms and senses. *WebNet J.: Internet Tech. Appl. Issues. 2*, 3, 54–61.

Received May 2001; revised December 2002 and February 2003; accepted July 2003