Form and Function: The Impact of Query Term and Operator Usage on Web Search Results

Wendy Lucas
Computer Information Systems Department, Bentley College, Waltham, MA 02452-4705.
E-mail: wlucas@bentley.edu

Heikki Topi
Computer Information Systems Department, Bentley College, Waltham, MA 02452-4705.
E-mail: htopi@bentley.edu

Conventional wisdom holds that queries to information retrieval systems will yield more relevant results if they contain multiple topic-related terms and use Boolean and phrase operators to enhance interpretation. Although studies have shown that the users of Web-based search engines typically enter short, term-based queries and rarely use search operators, little information exists concerning the effects of term and operator usage on the relevancy of search results. In this study, search engine users formulated queries on eight search topics. Each query was submitted to the user-specified search engine, and relevancy ratings for the retrieved pages were assigned. Expert-formulated queries were also submitted and provided a basis for comparing relevancy ratings across search engines. Data analysis based on our research model of the term and operator factors affecting relevancy was then conducted. The results show that the difference in the number of terms between expert and nonexpert searches, the percentage of matching terms between those searches, and the erroneous use of nonsupported operators in nonexpert searches explain most of the variation in the relevancy of search results. These findings highlight the need for designing search engine interfaces that provide greater support in the areas of term selection and operator usage.

Introduction

Buried within the pages of many search engine sites lie detailed instructions on improving the effectiveness of search results. Tips range from the general, such as "Be specific," to the specific, such as "Use the (+) sign to indicate must-have words." Yet search engine users typically ignore these rules and tips (Pollock & Hockley, 1997), believing they are intended for experienced searchers. Analyses of search engine transaction logs, including Jansen (2000) and Silverstein (Silverstein, Henzinger, Marais, & Moricz, 1999), have found that the majority of searchers enter term-based queries containing approximately two terms and rarely use operators. Search experts do use more terms and advanced search operators than the average Web user (Hölscher & Strube, 2000). They use fewer terms, however, than are typically found in queries to traditional information retrieval (IR) systems, which are three to seven times longer than an average Web query (Jansen, Spink, & Saracevic, 2000). Of the over one million queries to the Excite search engine analyzed in Spink, Wolfram, Jansen, and Saracevic (in press), less than 5% used any Boolean operators, and those that did often contained mistakes. This is not surprising, given that the rules governing operator usage vary from one search engine to another and are often difficult to find.

Although transaction logs provide descriptive statistics about the composition of users' queries, the intent of those queries, and hence the relevance of the documents they retrieve, is impossible to determine. In IR systems, ranking algorithms are commonly based on the match between query terms and the terms used to index the documents (see Gudivada, Radhavan, Grosky, & Kasanagottu, 1997; Harman, 1992). Web searches are different from their traditional counterparts in that they are conducted over largely unstructured data of uneven quality. This decreases the effectiveness of standard IR retrieval models for finding relevant pages. Search engine ranking algorithms attempt to factor in page quality by including measures based on the graphical structure of the Web, such as the number of links leading to and from a page (Brin & Page, 1998; Kleinberg, 1999), and on the popularity of a page, as determined by the number of visitors it receives and the amount of time they spend there. In light of the significant differences between well-defined database environments and the Web, one can-

DOI: 10.1002/asi.10013

not assume that the rules governing query composition that are used to great effect in the former environment can be applied equally as well to the latter.

The objective of this study is to determine the impact of query operators and term selection on the relevancy of Web search results. This knowledge is essential for the development of search interfaces that support the query formation process. Eighty-seven survey respondents formed queries on eight topics for use on a search engine of their choice. Each query was subsequently submitted to the specified search engine. A query formulated by search experts on each of these topics was also submitted to the eight preferred search engines of the respondents. Relevancy ratings were assigned to the first 10 documents retrieved. The query with the highest rating for each topic by search engine was then identified as the "expert query," regardless of whether it was formulated by a search expert or a survey respondent. Data analysis was then performed for determining the ways in which all other queries differed from the expert ones in their use of terms and operators, and for measuring the significance of those differences on the relevancy of search results.

The next section of this article describes studies related to the work presented here. This is followed by a description of our research model and the methodology of our study. Results are then presented and discussed. The article concludes with directions for future research.

Related Studies

Studies related to this work fall into three categories. The first includes evaluations of search engine performance. Their importance here stems from the fact that the effects of query terms and operators on the relevance of Web search results cannot be evaluated in isolation from the effectiveness of the individual search engines. The second category focuses on the use of terms and operators by search engine users. These studies are primarily limited to evaluations of user logs. A third category that includes user interactions with both traditional IR systems as well as Web search engines provides additional information on term selection and operator usage.

Information retrieval (IR) system performance is typically measured as its precision at various recall levels (Gudivada et al., 1997), where *precision* is defined as the ratio of the number of relevant documents retrieved to the total number of documents retrieved, and *recall* is the ratio of the number of relevant documents retrieved to the total number of relevant documents in the collection. Studies of search engine performance based on this metric have shown wide variations in the abilities of popular search engines to retrieve relevant pages (Chu & Rosenthal, 1996; Gordon & Pathak, 1999; Leighton & Srivastava, 1999). These variations will affect the optimal results one could achieve using a particular search engine, and therefore, must be factored in when comparing the effectiveness of queries across a variety of engines.

As already noted, analyses of search engine transaction logs have shown that most queries contain few terms and use a limited number of operators. A study of approximately one billion queries contained in the AltaVista query log (Silverstein et al., 1999) found that 72.4% had two or fewer terms, 79.6% contained no Boolean operators, and 15% were empty. In a study of 51,473 queries posed by 18,113 users of Excite (Jansen et al., 2000), it was found that the queries contained 2.21 terms on average, and that only 6% of the users made use of any Boolean capabilities. In Jansen (2000), 15 queries from this same transaction log were selected for further analysis. Criteria for query selection were based on studies describing the composition of typical queries to Web search engines. The selected queries were submitted to five search engines in their simplest form (i.e., with no advanced operators). They were then modified using the operators supported by each of the search engines and resubmitted. Overlap rates between the results returned by the simple queries and those returned by the advanced queries were calculated and used as the basis for comparison. Findings indicated that increasing query complexity by adding advanced operators had little effect, with an overlap of over 70% between the simple and complex query results. The relevancies of the retrieved documents were not considered in the analysis.

Studies investigating searching behavior have more frequently focused on user interaction with traditional IR systems rather than with Web search engines. Spink and Saracevic (1997) observed user interaction with the on-line database DIALOG and analyzed the retrieval effectiveness of search terms from five different sources. They found that 81% of terms from the question statement itself contributed to the retrieval of relevant items, as opposed to 71% of terms from relevance feedback, 52% from user interaction, 46% from a thesaurus, and 46% from a search intermediary.

A study of advanced humanities scholars using DIALOG over a 2-year period is described in Siegfried (Siegfried, Bates, & Wild, 1993). All were given a full day of DIALOG training before beginning their searches. Only 37.5% of their queries contained two or more words, and more than one-quarter of the scholars never used a Boolean OR, indicating that the 1 day of training was not enough to turn them into experienced searchers.

Hsieh-Yee (1993) compared the use of search terms and tactics by novice and expert searchers of bibliographic database systems to show that one must be knowledgeable about the search topic in addition to being technically competent to be successful in information retrieval. There was little variation in the number of terms used by novice versus expert searchers, suggesting the importance of term content versus term quantity. In Saracevic (Saracevic, Kantor, Chamis, & Trivison, 1988), the overlap in term selection among different search experts searching for the same question was found to be relatively low, with a mean of 0.27.

Hölscher and Strube (2000) also compared the searching habits of novice and expert Web searchers in a two-part study and found that successful Web searches rest on a

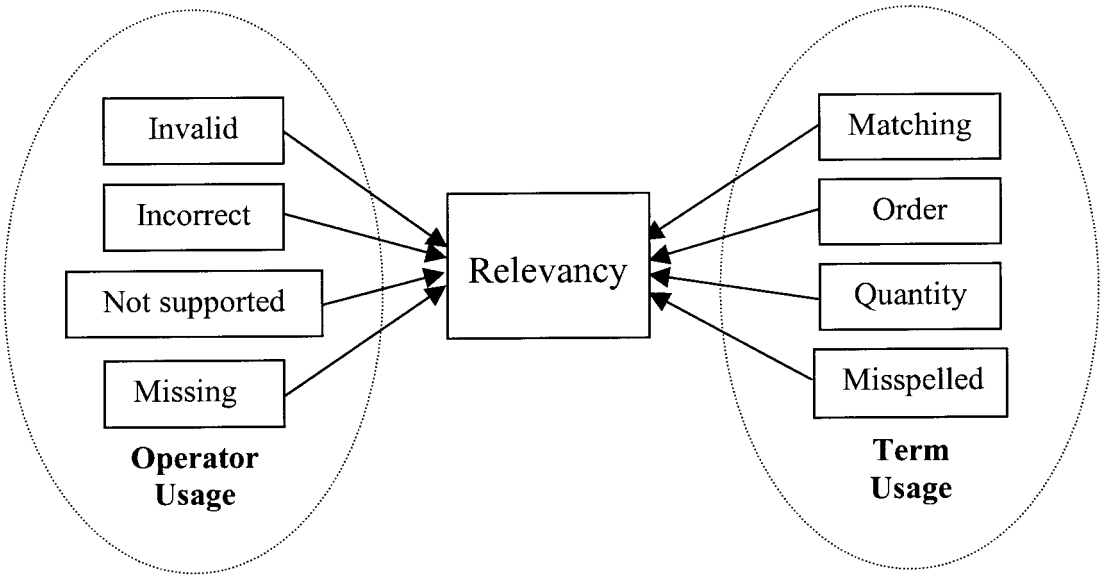

FIG. 1. Model of factors affecting relevancy of query results.

combination of experience and domain knowledge. In the first study, their 12 search experts used an average of 3.64 words per query, as opposed to an average of 1.66 words per query by users of Fireball, a German search engine from which 16 million queries were analyzed. They found that experts made use of advanced search options, including Boolean operators, modifiers, and phrases, more frequently than novices. The one exception was the (+) operator, which was equally popular among the two user groups.

In their second experiment involving 24 participants, the 12 expert searchers' queries contained an average of 2.61 words, while those of the novices was only slightly lower, with an average of 2.32 words per query. The four experts with greater domain knowledge, however, used only 1.97 words per query, as opposed to the 2.96 words per query used by the four search experts with little domain knowledge. Furthermore, the four novice searchers with high domain knowledge compensated for their lack of query formatting skills with greater verbal creativity. These findings support the concept that it is the terms that are chosen for use in queries, rather than their absolute number, that is significant.

Research Model

Figure 1 shows the research model that provides the basis for this study. It divides the query-specific factors affecting the relevancy of search results into two categories: operator usage and term usage. In the first of these categories, invalid operators include symbols, such as commas and semicolons, which have no meaning in queries but appear there nonetheless. Incorrectly used operators are valid but do not appear in the correct position within a query, such as before or between search terms. "Not supported" refers to the use of an otherwise valid operator that is not supported by the search engine to which the query is submitted. Finally, missing operators are identified as those that would appear before a particular term if that query were optimally formulated, but fail to appear there. An example is the absence of the Boolean NOT operator in the case where a term is to be excluded from rather than included in all retrieved documents.

In the term usage category, "matching," "order," and "quantity" factors are defined in terms of their similarities to an optimally formed query. "Matching" refers to the terms a query has in common with the optimal query for the same search topic. The order factor indicates if the terms in a query appear in the same order as those same terms in the optimal query. "Quantity" is the difference between the number of terms in the optimal query and the number in the query being evaluated. "Misspelled" corresponds to all terms in a query that have been spelled incorrectly.

The research model underlies the design of this study and the analysis of the collected data, which are described in the following sections.

Methodology

Data Collection

Subjects for this study were recruited via e-mail from students at Bentley College during the fall semester of 2000. Table 1 contains background information and Internet-related experience for the 87 participants.

Data was collected via a Web-based survey, enabling the subjects to choose the location and time best suited to their needs for participating. Each subject was asked to form queries on the eight topics shown in Figure 2. Optimal query formation would require some knowledge of conjunction,

TABLE 1. Demographic characteristics of the study subjects.

Age	Mean	Min	Max			
	25.75	19	48			
Gender	Male	Female				
	65.5%	34.5%				
	<6 months	6–12 months	1–3 years	3–5 years	>5 years	
Internet experience	0.0%	0.0%	5.7%	62.1%	32.2%	
Search engine exp.	0.0%	2.3%	18.4%	60.9%	18.4%	
Search engine skill	1 (novice)	2	3	4	5 (expert)	Mean
	0.0%	0.0%	24.1%	50.6%	25.3%	4.01

disjunction, negation, phrases, capitalization, stemming, and wildcards.

Subjects each specified a single preferred search engine and then entered queries to scrolling text boxes on the survey form, rather than to the search engine itself. This method was chosen to reduce cognitive load, and hence facilitate problem solving (Cooper, 1998), by focusing the respondents' attention on the search topics while they were forming queries without their having to switch views to an often-cluttered search engine interface. The search engines named by participants were AltaVista, AOL, Excite, Go, Google, iWon, Lycos, and Yahoo!.

Five expert searchers also formed queries on the eight search topics. These queries were then translated in conformance with the rules on operator usage and query syntax for seven of the search engines that were preferred by survey participants. Table 2 summarizes these rules for the basic search features of each of the search engines. It was compiled from the help pages of the search engines and information gleaned from our experiences using them. The use of quotation marks for denoting phrases is supported by all of the engines, so is not included in the table.

The appendix shows the translated queries for each search engine. Because Yahoo! uses Google for its Web page search results, the same translated queries were used for both. The expert and participant queries were subsequently submitted to the specified search engines for each search topic. Table 3 shows the number of queries by search engine. In all cases, Web search results, as opposed to directory listings, were then analyzed, as described next.

Query Analysis

For each search topic, all of the queries formulated by experts and survey subjects were run and analyzed during a 1-day period. This helped minimize differences in search results due to updating of the search engines' indices. First, the authors of this study ran the expert queries specified for each of the search engines. In judging the relevancy of pages retrieved in response to these queries, a cutoff value of 10 was used because links that appear on the first page of search results are the most likely to be viewed (Jansen et al., 2000; Silverstein et al., 1999). Understanding the impact of term and operator usage on the pages accessed from these links is key to improving the likelihood that the pages viewed by searchers will be relevant.

The retrieved pages were independently judged using a four-category ordinal scale for relevancy, support for which

Topic	Expert Query
1. How to make pudding.	1. `pudding AND recipe*`
2. Information on colleges located in Germany.	2. `Germany AND (college* OR universit*)`
3. Who said: The business of America is business.	3. `"the business of America is business" AND quot*`
4. Information about cowboys, but not the Dallas Cowboys.	4. `cowboy* AND NOT(dallas)`
5. Books written by Arthur Gittleman.	5. `"Arthur Gittleman" AND book*`
6. Information on the effects of caffeine on the heart.	6. `caffeine AND heart`
7. How to make cookies with peanuts but not peanut butter.	7. `cookie* AND peanut* AND recipe AND NOT("peanut butter")`
8. What John Silber does at Boston University.	8. `"John Silber" AND ("Boston University" OR BU)`

FIG. 2. Search topics and expert queries.

TABLE 2. Search engine support using basic search.

Search engine	Boolean AND OR NOT	(+) and (−) operators	Stemming	Capitals
AltaVista	Not supported	Supported	Supports wildcard (*)	Case sensitive
AOL	Supported Case insensitive	(+) Not required: defaults to matching all terms (−) Not supported	Supports wildcard (*)	Case insensitive
Excite	Supported Case insensitive	Supported Cannot leave space after operator	Not supported.	Case insensitive
Go/Infoseek	Supported Case insensitive	Supported Cannot leave space after operator	Supported	Case sensitive
Google/Yahoo!	Automatic AND. Supports OR No support for NOT Case sensitive	(+) not required: defaults to matching all terms Use (+) for stop words only (−) Supported Cannot leave space after operator	Not supported.	Case insensitive
iWon	Supported Case sensitive	Supported	Supports wildcard (*)	Case sensitive
Lycos	Supported Case sensitive	Supported Cannot leave space after operator	Supported	Case insensitive

is provided in Greisdorf (2001). A relevancy score of 3 corresponding to *highly relevant*, 2 corresponding to *relevant*, 1 corresponding to *somewhat relevant*, or 0 corresponding to *irrelevant* was assigned to the corresponding link. Prior to making these judgments, the criteria related to each of the relevancy scores for the search topic under consideration had been independently defined and mutually agreed upon. If duplicate and/or irretrievable pages were included in the first 10 links retrieved, then additional pages were analyzed until either a total of 10 was reached or there were no more links available. There were also cases where the total number of evaluated links was less than 10 due to fewer than that number being retrieved in response to a query.

The results of the independent evaluation of the retrieved links were then reviewed by both authors together. In the rare cases where the relevancy judgments were not in agreement, the Web pages were reviewed and a consensus was reached. The relevancy criteria were also reviewed, and adjustments were made to clarify their intent if necessary. The final, mutually agreed upon criteria for all of the search topics appear in the appendix.

These criteria were then used for evaluating responses to the queries formed by the survey subjects. Three evaluators, including the authors, participated in this process. To ensure consistency, the same person evaluated all of the pages retrieved for a particular search topic. Results of running a total of 645 queries covering the eight search topics were analyzed following this approach and are presented and discussed next.

TABLE 3. Number of queries by search engine.

Search engine	Number of queries	Percentage of total
AltaVista	61	9.5%
AOL	8	1.2%
Excite	16	2.5%
Go	8	1.2%
Google/Yahoo	508	78.8%
iWon	12	1.9%
Lycos	32	5.0%
Total	645	100.0%

Results

The purpose of this section is to present the empirical results of this study, with the detailed discussion and interpretation of results deferred until the following section. Descriptive statistics regarding the usage of operators and search terms are presented first. These are followed by the relevancy scores achieved by the study subjects and the expert searchers. We then present a correlation matrix for the variables included in our research model (see Fig. 1), succeeded by the results of the regression analysis, including both the full multiple regression and the step-wise hierarchical multiple regression.

Table 4 includes the basic descriptive statistics regarding the use of operators and search terms by research subjects for each of the individual search topics and for all of the queries taken together. First are the *Number* and the *Percentage of queries using operators* (including (+), (−), AND, OR, NOT, wildcards, and phrasing with quotes), followed by the same calculations for when phrasing with quotes is excluded from the set of operators. Also reported are the *Average total number of all operators* (including quotes) per query by the subjects and by the experts. Term usage statistics include the *Average number of terms used* for each of the search topics by the subjects and by the experts. In addition, the table includes the average number of terms used by the subjects that are the same as the terms used in the expert query (matching terms), and the corresponding percentage figure.

The percentage of queries in which subjects used operators varied widely, depending on the search topic. If quotes used for phrasing are categorized as operators, the percentages varied from 28.4% (Gittleman) to 63.3% (Business of

TABLE 4. Use of operators and terms by subjects.

Identifier	Topic 1 Pudding	Topic 2 German Colleges	Topic 3 Business of America	Topic 4 Cowboys	Topic 5 Gittleman	Topic 6 Caffeine	Topic 7 Cookies	Topic 8 Silber	Total
Total number of queries	87	86	79	79	81	80	75	78	645
Number of queries using operators	26	28	50	27	23	39	39	36	268
Percentage of queries using operators	29.9%	32.6%	63.3%	34.2%	28.4%	48.8%	52.0%	46.2%	41.4%
Number of queries using operators (excluding phrasing)	21	25	10	21	13	35	36	26	187
Percentage of queries using operators (excluding phrasing)	24.1%	29.1%	12.7%	26.6%	16.0%	43.8%	48.0%	33.3%	29.0%
Average total number of all operators used	0.42	0.52	0.78	0.53	0.43	0.65	0.92	0.86	0.63
Average total number of all operators used by experts	1.71	3.14	2.86	1.71	2	1.29	5.14	3.86	2.66
Average number of terms used	1.93	2.31	4.78	1.91	2.21	2.7	3.07	3.35	2.76
Average number of terms matching expert query	1.38	1.47	4.18	0.48	1.96	1.94	1.45	3.15	2
Percentage of terms matching the expert query	71.5%	63.6%	87.4%	25.1%	88.7%	71.9%	47.2%	94.0%	57.90%
Average number of terms used by experts	2.14	2.71	7.14	2	2.71	2.14	4.14	4.714	3.43

America). These relatively high numbers change, however, if quotes are excluded from the set of operators, in whch case Search topic 3 (Business of America) moves to the lowest position (operators used in 12.7% of the queries) and Search topic 7 (Cookies) has the highest percentage (operators used in 48% of the queries). The same sort of variation by search topic applies to the use of terms: the lowest average number of terms both for the expert queries and the subject queries was approximately 2 (Search topic 4, Cowboys), while the highest number was 4.78 for subject and 7.14 for the experts (both for Search topic 3, Business of America). Term selection accuracy, defined as the percentage of terms used by subjects that matched those used in the expert queries, also varied widely by search topic from 47.2% (Cookies) to 88.7% (Gittleman).

In Table 5, we provide a more detailed analysis of the usage of operators by the research subjects. This table reports the total number of operators used by operator type [(+), (−), AND, OR, NOT, wildcards, and phrasing with queries] for each search topic, *not* the number of queries in which the operators are used. The lower half of the table reports the number of operators that were missing from study subjects' queries, categorized by search topic and operator type. An operator was considered to be missing

TABLE 5. Detailed analysis of operator usage.

Identifier	Topic 1 Pudding	Topic 2 German Colleges	Topic 3 Business of America	Topic 4 Cowboys	Topic 5 Gittleman	Topic 6 Caffeine	Topic 7 Cookies	Topic 8 Silber	Total
Total no. of queries	87	86	79	79	81	80	75	78	645
No. of operators used by subjects									
+	20	22	15	8	17	21	30	23	156
−	0	0	2	7	0	0	7	0	16
AND	6	12	4	3	4	25	15	14	83
OR	2	1	0	0	0	0	0	0	3
NOT	0	0	0	10	0	0	7	0	17
wildcards	0	0	0	1	0	0	0	0	1
phrasing with quotes	9	10	40	13	14	6	10	30	132
No. of operators missed by subjects									
+	6	11	4	0	15	3	20	19	78
−	0	0	0	1	0	0	2	0	3
AND	6	4	1	1	1	9	5	2	29
OR	0	0	0	0	0	0	0	0	0
NOT	0	0	0	1	0	0	1	0	2
wildcards	6	5	3	11	5	0	25	0	55
phrasing with quotes	0	0	15	2	45	1	5	82	150

TABLE 6. Number of terms used by subjects.

	Number of subjects using a specific number of terms									
	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8		
Identifier	Pudding	German Colleges	Business of America	Cowboys	Gittleman	Caffeine	Cookies	Silber	Total	Total %
Total no. of queries	87	86	79	79	81	80	75	78	645	
No. of terms										
1	22	4	7	41	9	13	5	0	101	15.7%
2	54	58	12	16	50	33	25	21	269	41.7%
3	7	22	2	16	18	22	21	16	124	19.2%
4	3	1	3	2	4	5	12	37	67	10.4%
5	1	0	11	3	0	2	8	3	28	4.3%
6	0	0	37	0	0	4	4	0	45	7.0%
7	0	0	4	1	0	0	0	0	5	0.8%
8	0	0	3	0	0	0	0	1	4	0.6%
>8	0	1	0	0	0	1	0	0	2	0.3%

only if the subject's query included the term(s) with which the operator was associated in the expert query. For example, the AltaVista expert query for Search topic 1 (Pudding) is *+pudding +recipe**. If a subject's query were *pudding recipe*, it would have been evaluated as having all three operators missing because the corresponding terms were included; if, instead, the subject's query had been *pudding*, it would have only one missing operator [(+) associated with *pudding*]. Table 6 provides a simple categorization of the numbers of search terms used in the research subjects' queries by search topic.

The detailed analysis of operator usage reveals that the (+) operator was by far the most widely used (altogether 156 instances), followed by the only other significant operator, AND (83 instances). In addition, subjects used double quotes for phrasing in 132 instances. The other operators were barely used, varying from 1 instance of wildcard usage to 17 instances of the use of the NOT operator. Operator usage is, of course, strongly dependent on the information need and on the terms a subject has included in the query. Table 5 reveals that, for these particular queries, the operators that were most commonly used were also the most commonly missing. The only exception is with wildcards, whose low number is truly an indication that subjects frequently ignore the opportunity to improve queries with their use. It is important to note, however, that in many case, and most particularly with (−) and NOT, an operator was not considered missing because the related term was missing, as previously described.

Table 6 supports the findings of earlier research that subjects use a relatively small number of terms in their searches. In the entire population, almost 60% of the queries had only one or two terms, and more than 75% had only one, two, or three terms. Of even greater importance is that these percentages depend strongly on the search topic: with Search topic 1 (Pudding), 87% of the queries used only one or two terms, whereas with Search topic 3 (Business of America), 24% used one or two terms and 56% used more than five terms.

Relevance is the focus of Table 7. Three different relevance figures are reported for each of the queries: the average relevance of the expert queries, which is used as the point of comparison, the average relevance of the queries written by the study subjects; and the average standardized (subject) relevance for the same search topic/search engine combination. Please note that the average standardized relevance is not the same as the average subject relevance divided by the average expert relevance because this ratio ignores differences between the optimal performances achieved with each search engine. As mentioned earlier in the Methodology section, the selection of the expert query and the evaluation of subject queries for a particular search topic were done within one calendar day to minimize the effect of the dynamic nature of the query results on relevancy comparisons.

The relevance levels achieved also varied greatly, depending on the search topic for both expert searchers and subjects. Experts achieved very high levels of relevance

TABLE 7. Average expert and subject relevancies by query type.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	
Identifier	Pudding	German Colleges	Business of America	Cowboys	Gittleman	Caffeine	Cookies	Silber	Total
Total number of queries	87	86	79	79	81	80	75	78	645
Expert relevance	2.69	2.47	2.54	2.26	1.44	2.23	1.39	2.9	1.93
Subject relevance	1.54	1.88	0.38	0.99	0.83	1.68	0.65	2.09	1.26
Standardized subject relevance	0.53	0.64	0.15	0.4	0.59	0.7	0.3	0.7	0.5

TABLE 8. First-order correlations between research variables.

	1	2	3	4	5	6	7	8	9	10	11
1 Average standardized relevance											
2 Total No. of missing operators	0.059										
3 No. of invalid operators	0.014	0.050									
4 No. of nonsupported ANDs	**0.127**	0.037	−0.069								
5 No. of nonsupported '+'s	**0.089**	−0.062	0.015	−0.076							
6 No. of other nonsupported operators	**−0.197**	−0.073	0.012	0.024	−0.020						
7 No. of incorrectly used operators	−0.059	−0.013	−0.044	−0.047	−0.041	−0.025					
8 No. of correctly used operators	0.022	0.037	−0.006	**−0.085**	0.046	0.036	−0.009				
9 Difference between numbers of terms	**−0.417**	−0.027	−0.023	−0.064	−0.009	0.051	**0.091**	−0.050			
10 Percentage of terms matching	**0.399**	**0.176**	**0.078**	**0.101**	**0.086**	−0.006	0.058	**0.252**	**−0.314**		
11 Order of terms	0.011	−0.042	**0.171**	0.037	**0.173**	−0.012	−0.031	0.049	**−0.156**	0.047	
12 Number of misspelled terms	**−0.107**	0.018	−0.016	0.024	**0.292**	−0.024	**0.086**	0.049	0.044	−0.030	0.011

Note: **bold** typeface indicates a statistically significant correlation at the 0.05 level.

with two of the search topics—1 (Pudding; 2.69) and 8 (Silber; 2.90)—but even the expert search performance was poor for two fo the search topics, 7 (Cookies; 1.39) and 5 (Gittleman; 1.44). The subjects were able to achieve the highest performance level with Search topic 8 (Silber; 2.09). Their worst results were with a search topic for which experts were relatively successful, Search topic 3 (Business of America; 0.38). The average standardized relevance of the subject queries was, therefore, the lowest with Search topic 3 (15%). It was at its highest levels with Search topics 6 (Caffeine) and 8 (Silber), both 70%.

Table 8 presents the first-order Pearson correlations between the main research variables that are included in the research model (see Fig. 1). The variables include the average standardized relevance (as described above), seven variables related to the use of operators, and four variables related to the use of terms. The variables related to the use of operators are the total number of missing operators (including missing phrasing with quotes), the total number of invalid operators (such as commas, semicolons, question marks, and arrows, which are not actually operators in these query languages), the total number of nonsupported ANDs (i.e., AND operators that were used with search engines that do not support them), the total number of nonsupported (+) operators [i.e., (+) operators that were used with search engines that do not support them], and the total number of other nonsupported but valid operators. The variables related to term usage are the absolute difference between the number of query terms in the subject query and the corresponding expert query, the percentage of subject query terms matching the terms in the expert query, a binary variable telling whether the order of the query terms in the subject query was the same as the order of the terms in the expert query (1 = incorrect), and the number of misspelled query terms.

As Table 8 indicates, there is a statistically significant (at the 0.05 level) correlation between the main performance variable, *Average standard relevance*, and six of the independent variables. Clearly, the strongest of these are *Difference between numbers of terms* (−0.42) and *Percentage of terms matching* (0.40). It is also worth noting the high correlations between *Number of nonsupported* (+) *operators* and *Number of misspelled terms* (0.29) and between *Number of correctly used operators* and *Percentage of terms matching* (0.25). *Difference between numbers of terms* is negatively correlated with *Percentage of terms matching* (−0.31) and *Order of terms* (−0.16).

The regression analyses are presented in Tables 9 and 10. In Table 9 are the results of the regression model in which all of the independent variables were entered into the equation simulatneously. Table 10 shows the first four steps of a step-wise hierarchial multiple regression based on the same set of independent variables. The dependent variable in both of these models was *Average standardized relevance*, which is used as a surrogate for the subject's query writing performance. The independent variables are the same as those described above in the context of the correlation matirx.

The overall analysis in Table 9 shows that the model explains a respectable 32% of the variance in the dependent variable and is highly significant. The most significant independent variables in the full model are *Difference between numbers of terms*, $t(629) = -8.85$, $p < 0.001$, *Percentage of terms matching,* $t(629) = 7.99$, $p < 0.001$, *Number of other nonsupported operator*s, $t(629) = -5.51$, $p < 0.001$, and *Number of misspelled terms*, $t(629) = -3.36$, $p = 0.001$. For all of these variables, the sign of the regression coefficient is as expected: the only independent variable in this group with a positive association with the dependent variable is *Percentage of terms matching*. Three additional variables that are also significant in the full

TABLE 9. Results of the full multiple regression.

Dependent variable: Average standardized relevance

$R = .564 \quad R^2 = .318 \quad$ Adjusted $R^2 = .306$

$F(11, 630) = 26.752 \quad p < .00000$ Std. Error of estimate: .289

	BETA	St. Err. of BETA	B	St. Err. of B	$t(629)$	p-level
Intercept			0.430	0.032	13.412	0.000
Total number of missing operators	−0.011	0.034	−0.004	0.013	−0.337	0.736
Number of invalid operators	0.000	0.034	0.000	0.031	−0.010	0.992
Number of nonsupported ANDs	0.090	0.034	0.106	0.040	2.668	0.008++
Number of nonsupported '+'s	0.111	0.036	0.073	0.023	3.121	0.002++
Number of other nonsupported operators	−0.183	0.033	−0.418	0.076	−5.512	0.000++
Number of incorrectly used operators	−0.036	0.034	−0.067	0.062	−1.072	0.284
Number of correctly used operators	−0.051	0.034	−0.023	0.016	−1.473	0.141
Difference between numbers of terms	−0.313	0.035	−0.080	0.009	−8.847	0.000++
Percentage of terms matching	0.297	0.037	0.339	0.042	7.986	0.000++
Order of terms	−0.074	0.034	−0.056	0.026	−2.146	0.032+
Number of misspelled terms	−0.117	0.035	−0.268	0.080	−3.356	0.001++

Note: ++ indicates a p-value <.01; + indicates a p-value <.05

model had unexpected signs: *Number of nonsupported (+) operators*, $t(629) = 3.12$, $p = 0.002$, *Number of nonsupported AND operators,* $t(629) = 2.67$, $p = 0.008$, and *Order of terms*, $t(629) = -2.15$, $p = 0.032$.

The exploratory step-wise hierarchical regression shown in Table 10 is consistent with the full model. The order in which the step-wise regression enters the variables is as follows: *Difference between numbers of terms* (negative coefficient), *Percentage of terms matching* (positive coefficient), *Number of other nonsupported operators* (negative coefficient), and *Number of misspelled terms* (negative coefficient). These four variables together explain 29.3% of the variance, which is almost the entire explained variance by the full model.

Discussion

In this section, we will discuss the meaning and the interpretation of our results, with an emphasis on the most important empirical findings of the study. One of the key findings is an obvious one that is often overlooked in prior research: The information request on which a query is based has a strong impact on both the numbers of operators and terms used to formulate a query. The average number of all operators used by the subjects was low (0.63), and only 41.4% of the queries used any operators [this figure is higher than those reported in Jansen (2000) and Silverstein (1999) due to our inclusion of (+) and (−) operators in addition to Booleans]. Both of these numbers varied widely, however, depending on the search topic. Simple information request such as "How to make pudding?" (Search topic 1) and "Books written by Arthur Gittleman" (Search topic 5) had very low values for *Average total number of all operators used* (0.42 and 0.43, respectively). In addition, less than 30% of the queries written for these search topics contained operators. By contrast, the clearly more complex search topic of "How to make cookies with peanuts but not peanut butter" (Search topic 7) generated a significantly higher average number of operators (0.92) and more than 50% of the corresponding queries including operators. These differences were even more pronounced with the expert queries: average operator usage was lowest (1.3) for "Information on the effects of caffeine on the heart" (Search topic 6), and highest (5.14) for the same search topic as for the research subjects, namely Search topic 7. Overall, experts used more operators than the research subjects, which agrees with the findings reported in Hölscher (2000).

The experts also used more terms than the subjects, although the differences were not as pronounced as with operator usage (overall 3.43 for the experts and 2.76 for the subjects). This is in keeping with Hölscher (2000) and Hsieh-Yee (1993), who also found that the number of terms did not vary greatly between novice and expert searchers. The difference in the number of terms used by search topic, however, was far more obvious. The search topic generating queries with the lowest number of terms was "Information about cowboys, but not the Dallas Cowboys" (Search topic 4; average 2.00 for the experts and 1.91 for the subjects), whereas the highest number of terms was linked to the search for "Who said: The business of America is business?" (Search topic 3; average 7.14 for the experts and 4.78 for the subjects). This same phenomenon is demonstrated by the more detailed analysis shown in Table 6. Thus, our results indicate that broadly generalized statements regarding the number of terms or operators used by searchers are easily misleading. Evaluating these characteristics in the context of a specific information request significantly improves our understanding of their true meaning.

For our specific set of search topics, the terms used by the research subjects matched those in the expert queries surprisingly well in six of the eight cases. The exceptions were for Search topics 4 (Cowboys) and 7 (Cookies), where the sub-

TABLE 10. Results of the step-wise hierarchical multiple regression.

Dependent variable: Average standardized relevance

Step 1. Insert Difference between numbers of terms

$R = .417 \quad R^2 = .174 \quad$ Adjusted $R^2 = .172$

$F(1, 640) = 134.42 \quad p < 0.0000$ Std. Error of estimate: .315

	BETA	St. Err. of BETA	B	St.Err. of B	$t(639)$	p-level
Intercept			0.412	0.031	13.306	0.000
Difference between numbers of terms	−0.417	0.036	−0.107	0.009	−11.594	0.000

Step 2. Insert Percentage of terms matching

$R = .503 \quad R^2 = .253 \quad$ Adjusted $R^2 = .251$

$F(2.639) = 108.3 \quad p < 0.0000$ Std. Error of estimate: .300

	BETA	St. Err. of BETA	B	St. Err. of B	$t(638)$	p-level
Intercept			0.407	0.032	12.869	0.000
Difference between numbers of terms	−0.323	0.036	−0.083	0.009	−8.985	0.000
Percentage of terms matching	0.297	0.036	0.339	0.041	8.251	0.000

Step 3. Insert Number of other nonsupported operators

$R = .534 \quad R^2 = .285 \quad$ Adjusted $R^2 = .282$

$F(3, 638) = 84.809 \quad p < 0.0000$ Std. Error of estimate: .294

	BETA	St. Err. of BETA	B	St. Err. of B	$t(637)$	p-level
Intercept			0.412	0.031	13.306	0.000
Difference between numbers of terms	−0.314	0.035	−0.080	0.009	−8.887	0.000
Percentage of terms matching	0.299	0.035	0.341	0.040	8.482	0.000
Number of other nonsupported operators	−0.179	0.034	−0.410	0.077	−5.340	0.000

Step 4. Insert Number of misspelled terms

$R = .541 \quad R^2 = .293 \quad$ Adjusted $R^2 = .289$

$F(4, 637) = 66.003 \quad p < 0.0000$ Std. Error of estimate: .292

	BETA	St. Err. of BETA	B	St. Err. of B	$t(637)$	p-level
Intercept			0.417	0.031	13.504	0.000
Difference between numbers of terms	−0.310	0.035	−0.079	0.009	−8.823	0.000
Percentage of terms matching	0.297	0.035	0.339	0.040	8.476	0.000
Number of other nonsupported operators	−0.181	0.033	−0.416	0.077	−5.433	0.000
Number of misspelled terms	−0.089	0.033	−0.204	0.076	−2.671	0.008

jects' terms had a 64 and 89% match, respectively, with the expert terms (i.e., for Cowboys, 64% of the terms that subjects used were amongst the terms included in the expert query). Search topic 4 had a match of only 25%, and the match for Search topic 7 was also low at 47%. The reasons underlying these differences must be carefully analyzed, as highlighted by the Cowboy search topic. The final expert query on this topic for the Google search engine was +"*cowboy*" –"*Dallas*." The search term "cowboys," found in 52 out of a total of 69 queries, does not match the expert search term "cowboy," because Google does not perform word stemming.

The detailed analysis of the operators *used* and not used, or *missed*, by the subjects confirms that the (+) operator for inclusion and quotes for phrasing are clearly the most familiar and best understood. In addition, the AND operator was used for some search topics, i.e., topics 2, 6, 7, and 8, with topic 6 ("Information on the effects of caffeine on the heart") accounting for almost one-third of all AND operator usage. For those queries that required the use of either NOT or (−), it is important to note that the data here does not reveal all of those cases where the operator was missed. This is because an operator was considered missing only if the term with which it was associated in the expert query was present in the subject's query. Closer analysis of the data reveals that in queries where a logical NOT was part of the expert solution, either NOT or (−) was only included in the subjects' queries about 20% of the time. As noted in Cooper (1988) the use of Boolean operators is confusing, even after one has learned their correct interpretation.

The results related to the average relevancies for each of the search topics are revealing in many respects. Once again, for both expert and subject searches, there are significant differences between the various topics. The average relevance achieved by the experts varied from 1.39 (Search topic 7, Cookies) to 2.9 (Search topic 8, Silber). When evaluating these relevancies, it is important to remember that they are the average over all of the search engines using the 10 best results from each. Search topics 5 and 7 had

average expert relevancies that were below 2 (1.44 and 1.39, respectively), while all others were above that threshold. For the research subjects, the relevancies were clearly lower than for the experts, particularly for Search topics 3, 4, 5, and 7, in which average relevance was well below 1. The results obtained by the subjects were simply poor, and were far removed from the potential relevancies that could have been achieved with better queries. The subjects' overall average standardized relevance was 0.5 (50% of the expert results), and for three of the eight search topics (3—Business of America, 4—Cowboys, and 7—Cookies) it was 40% or below. The lowest was 15% for Search topic 3. Interestingly, Search topics 4 and 7 are the ones that require the use of the NOT operator.

The most important findings of this study are those related to the research model shown in Figure 1 and analyzed using the two different regression analyses presented in Tables 9 and 10. Our intention is to begin an exploration of the factors that have the most significant impact on subjects' performance in information retrieval tasks. Performance was operationalized with *Average standardized relevance*, the dependent variable in the regression model. Seven variables related to operator usage and four variables related to term usage served as independent variables. First, using the full model, we found that these variables taken together explain 31.8% of the variance in the dependent variable, which is both statistically significant and a relevant amount of variance. The statistical significances are, therefore, not just caused by the relatively large sample size. The significance of the overall model gives us a foundation on which to base any further analysis. This leads to our second conclusion, which is strongly supported by both analyses: in this context, **search term** selection and usage are much more important than the selection and usage of operators.

The two independent variables most strongly associated with the dependent variable are both term variables, one measuring the number of terms (as the absolute difference between the number of terms in a subject's query and a corresponding expert query) and the other measuring the number of subject terms that match terms in the corresponding expert query. These two variables together explain more than 25% of the variance in the dependent variable. In addition, the number of misspelled terms is also significant as a predictor of search performance. These findings suggest that in the training and support materials developed for search engines, strong attention must be paid to issues related to the selection and use of search terms, which are often ignored. The results of this study clearly indicate that the selection of search terms matters, and should be evaluated carefully when looking for the best possible search results. They also suggest that, if a searcher is not pleased with search results, even small changes may lead to significantly different—and potentially better—outcomes.

Among the four most important independent variables is only one operator variable, the *Number of other nonsupported operators*. It measures the number of NOTs, ORs, and (−) operators in contexts where they are not supported. This particular result and other observations based on our data suggest that the subjects have a tendency to use operators without considering the characteristics of a particular search engine, leading to negative consequences in most, but not all, cases. One of the surprising results of our study is that the use of nonsupported ANDs and nonsupported (+) operators was positively associated with search performance. This is an interesting finding that requires further analysis.

Conclusions and Future Work

We have presented an exploratory study that is, to the best of our knowledge, the first to test the relative strengths of term and operator usage on the relevancy of search results from Web search engines.

One of our main findings is the influence of the information request on both the number of operators and the number of terms used to formulate a query. Although prior studies have found that Web searchers typically use few terms and operators in their queries (Hölscher & Strube, 2000; Jansen et al., 2000; Silverstein et al., 1999), we have extended these works by showing that the search topic itself has a significant effect on the number of components that make up a query. Its inclusion in any analysis of term and operator usage is therefore essential for developing an understanding of both the query formation process and its impact on search results.

Our other findings grew out of our research model, which divides the factors affecting the relevancy of search results into the two categories of operator usage and term usage. Using this model as the basis for our analysis, we found that search term selection and usage are much more significant predictors of query performance than their operator counterparts. The first two independent variables most strongly associated with the dependent variable of average standardized relevance are the absolute difference in the number of terms between the expert and subject query, and the number of subject terms that match terms in the corresponding expert query. These findings extend the work of Jansen (2000), which indicated that increasing query complexity by adding advanced operators had little effect on query results.

Limitations of this study arose primarily from our desire to reduce the cognitive load on the research subjects, which led to an artificial environment for specifying queries. In addition, we did not have full control of that environment, and so are unable to know if the subjects sought help from search engine help pages or from others in formulating their queries.

An obvious direction for future work is in the area of search interface usability. Prior to the introduction and widespread use of Web search engines, users of IR systems were trained, dedicated searchers (Harman, 1994). This profile no longer fits the average user of Web search engines, as evidenced by the poor query formation skills of the participants in this study. The results from the analysis presented here will be applied to the design and development of a search interface that supports the query formation processes of today's search engine users. Further investigation will also be conducted into the factors affecting a searcher's choice of terms and use of operators.

Methodologies for improving users' understanding of the relationship between documents and query terms, such as visualizations of search activities and results, provide additional avenues for future research.

Acknowledgments

We are grateful to the expert searchers for their contributions to this study, and to the students at Bentley College who participated in our survey.

References

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. Paper presented at the 7th International World Wide Web Conference.

Chu, H., & Rosenthal, M. (1996, October 19–24). Search Engines for the World Wide Web: A comparative study and evaluation methodology. Paper presented at the ASIS 1996 Annual Conference.

Cooper, G. (1998). Research into cognitive load theory and instructional design at UNSW. University of New South Wales, Sydney, Australia. Available: http://www.arts.unsw.edu.au/education/CLT_NET_Aug_97.HTML [2001, March 14].

Cooper, W. (1988). Getting beyond Boole. Information Processing and Management, 24, 243–248.

Gordon, M., & Pathak, P. (1999). Finding information on the World Wide Web: The retrieval effectiveness of search engines. Information Processing & Management, 35(2), 141–180.

Greisdorf, H., & Spink, A. (2001). Median measure: An approach to IR systems evaluation. Information Processing and Management.

Gudivada, V.N., Raghavan, V.V., Grosky, W.I., & Kasanagottu, R. (1997). Information Retrieval on the World Wide Web. IEEE Internet Computing, 1(5), 58–68.

Harman, D. (1992). Ranking algorithms. In W.B. Frakes & R. Baeza-Yates (Eds.), Information retrieval: Data structures & algorithms (pp. 363–392). Upper Saddle River, NJ: Prentice Hall.

Harman, D. (1994). Overview of the Second Text REtrieval Conference (TREC-2), NIST SP 500-215, National Institute of Standards and Technology, Gaithersburg, MD.

Hölscher, C., & Strube, G. (2000). Web search behavior of Internet experts and newbies. The International Journal of Computer and Telecommunications Networking, 33(1–6), 337–346.

Hsieh-Yee, I. (1993). Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. Journal of the American Society for Information Science, 45(3), 161–174.

Jansen, B.J. (2000). The effect of query complexity on Web searching results. Information Research, 6(1).

Jansen, B.J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the Web. Information Processing and Managment, 36(2), 207–227.

Kleinberg, J.M. (1999). Authoritative sources in a hyperlinked environment. Journal of the ACM, 46(5), 604–632.

Leighton, H.V., & Srivastava, J. (1999). First 20 precision among World Wide Web search services (search engines). Journal of the American Society for Information Science, 50(10), 870–881.

Pollock, A., & Hockley, A. (1997). What's wrong with Internet searching. D-lib magazine, March.

Saracevic, T., Kantor, P., Chamis, A.Y., & Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. II. Users, questions, and effectiveness. III. Searchers, searches, and overlap. Journal of the American Society for Information Science, 39(3), 161–216.

Siegfried, S., Bates, M.J., & Wilde, D.N. (1993). A profile of end-user searching behavior by humanities scholars: The Getty online searching project report no. 2. Journal of the American Society for Information Science, 44(5), 273–291.

Silverstein, C., Henzinger, M., Marais, J., & Moricz, M. (1999). Analysis of a very large web search engine query log. SIGIR Forum, 33(1), 6–12.

Spink, A., & Saracevic, T. (1997). Interactive information retrieval: Sources and effectiveness of search terms during mediated online searching. Journal of the American Society for Information Science, 48(8), 741–761.

Spink, A., Wolfram, D., Jansen, B.J., & Saracevic, T. (in press). Searching the web: The public and their queries. Journal of the American Society for Information Science, 52(3).

Appendix: Expert Queries and Relevance Criteria

"Expert" queries are defined here as those that resulted in the highest relevancy ratings from the search engine, regardless of whether they were formed by an expert searcher or a survey responder.

1. **How to make pudding.**

   General expert search: pudding AND recipe*

   Expert search by search engine:

AltaVista:	+pudding +recipe*
AOL:	pudding AND recipes
Excite:	pudding AND recipes
Go/Infoseek:	pudding AND recipe*
Google/Yahoo!:	pudding recipe OR recipes
iWon:	pudding AND recipes
Lycos:	"pudding" and "ingredients"

   Relevancy Criteria:

   3: Contains at least two pudding recipes and/or links to pages with additional pudding recipes.
   2: Contains one pudding recipe.
   1: Contains the term pudding, perhaps as an ingredient in a recipe, but does not define how to make pudding.
   0: Either does not contain the term pudding or uses it in another context than that of food.

2. **Information on colleges located in Germany.**

   General expert search: Germany AND (college* OR universit*)

   Expert search by search engine:

AltaVista:	+colleges +germany
AOL:	Germany AND (college* OR universit*)
Excite:	Germany AND (college* OR universit*)
Go/Infoseek:	Germany AND universities
Google/Yahoo!:	colleges Germany
iWon:	Germany AND college* OR universit*
Lycos:	+Germany +universit*

   Relevancy Criteria:

   3: Contains or links to information on at least two colleges or universities in Germany or describes the higher education system in Germany.
   2: Contains or links to information on one college or university in Germany.

1: May mention Germany and colleges/universities, but not ones located in Germany, or mentions German colleges or universities but does not contain any information on them.
0: May contain information on Germany or on colleges/universities but not on both.

3. **Who said: *The business of America is business*.**

General expert search: "the business of America is business" AND quot*

Expert search by search engine:

AltaVista:	+"The business of America is business" +quot*
AOL:	"The business of America is business" AND quot*
Excite:	"The business of America is business" AND (quote OR quotation)
Go/Infoseek:	"The business of America is business" +quote
Google/Yahoo!:	"The business of America is business" quote OR quotation
iWon:	"The business of America is business" AND quot*
Lycos:	"The business of America is business"

Relevancy Criteria:
3: Explicitly states who said, "The business of America is business."
2: Contains a reference to both "The business of America is business" and its source but does not explicitly link him to the quotation.
1: Contains the phrase and some additional information about who said it but does not specify who that person is.
0: May or may not contain the phrase and does not mention any information about the source of the quotation.

4. **Information about cowboys, but not the Dallas Cowboys.**

General expert search: cowboy* AND NOT(dallas)

Expert search by search engine:

AltaVista:	cowboy* -dallas
AOL:	cowboys AND NOT dallas
Excite:	cowboy AND NOT dallas
Go/Infoseek:	cowboys -dallas
Google/Yahoo!:	+"cowboy" -"Dallas"
iWon:	cowboys AND NOT dallas
Lycos:	cowboy* -dallas

Relevancy Criteria:
3: Contains or links to a variety of information about cowboys and their interests, including ranches, horses, rodeos, etc.
2: Contains information on one topic relevant to cowboys and their interests.
1: Mentions cowboys but does not contain any information about them or relevant to their interests.
0: Either does not contain anything about cowboys or contains information about football (i.e., the Dallas Cowboys).

5. **Books written by Arthur Gittleman.**

General expert search: "Arthur Gittleman" AND book*

Expert search by search engine:

AltaVista:	+"Arthur Gittleman" +book*
AOL:	"Arthur Gittleman" AND books
Excite:	"Arthur Gittleman" AND books
Go/Infoseek:	Arthur Gittleman
Google/Yahoo!:	"Arthur Gittleman" books
iWon:	"Arthur Gittleman" AND books
Lycos:	+"Arthur Gittleman" +books

Relevancy Criteria:
3: Mentions the title of several books written by Arthur Gittleman.
2: Mentions the title of one book written by Arthur Gittleman.
1: Mentions Arthur Gittleman and a book or books, but not those written by him.
0: May contain information about books or mention an Arthur Gittleman who is not the author of any books, but not about both Gittleman and books.

6. **Information on the effects of caffeine on the heart.**

General expert search: caffeine AND heart

Expert search by search engine:

AltaVista:	caffeine and health
AOL:	caffeine AND heart
Excite:	caffeine AND heart
Go/Infoseek:	+caffeine +heart
Google/Yahoo!:	caffeine heart
iWon:	caffeine AND heart
Lycos:	+caffeine +heart

Relevancy Criteria:
3: Contains or links to extensive information on the effects of caffeine on the heart (such as an article, a caffeine/heart portal, or a description of personal experiences).
2: Contains or links to a limited amount of material related to the effects of caffeine on the heart.
1: Contains information about caffeine and about the heart, but does not relate one to the other.
0: May include the term caffeine and the term heart, but has no information about either, or may contain information about caffeine or the heart, but not about both.

7. **How to make cookies with peanuts but not peanut butter.**

General expert search: cookie* AND peanut* AND recipe AND NOT("peanut butter")

Expert search by search engine:

AltaVista:	+cookie* +recipe* +peanut* -"peanut butter"
AOL:	peanuts+cookies
Excite:	+recipe +cookies +peanuts -"peanut butter"
Go/Infoseek:	+cookie* +recipe* +peanut* -"peanut butter"
Google/Yahoo!:	"peanut cookies"
iWon:	cookie* AND recipe* AND peanut* AND NOT "peanut butter"
Lycos:	+cookie* +recipe* +peanut* -"peanut butter"

Relevancy Criteria:

3: Contains or directly links to two or more recipes for cookies containing peanuts but not peanut butter.
2: Contains or directly links to one recipe for cookies containing peanuts but not peanut butter.
1: Contains recipes for cookies and recipes containing peanuts, but not recipes for making cookies containing peanuts, or mentions cookies with peanuts but does not describe how to make them, or mentions both cookies and peanuts.
0: May contain a reference to either cookies or peanuts, but not both, or contains references to peanut butter.

8. **What John Silber does at Boston University.**

General expert search: "John Silber" AND ("Boston University" OR BU)

Expert search by search engine:

AltaVista:	+"John Silber" "Boston University" BU
AOL:	"John Silber" AND ("Boston University" OR BU)
Excite:	"John Silber" AND ("Boston University" OR BU)
Go/Infoseek:	"John Silber" AND "Boston University" OR BU
Google/Yahoo!:	"John Silber" "Boston University"
iWon:	"John Silber" AND "Boston University" OR BU
Lycos:	+"John Silber" +"Boston University"

Relevancy Criteria:

3: Explicitly states John Silber's position at Boston University.
2: Creates an association between John Silber and Boston University and implicitly makes it possible to infer his position there.
1: Contains information about John Silber and about Boston University, but does not relate one to the other.
0: May contain information about either John Silber or Boston University, but not about both.