

# The effect of query complexity on Web searching results

[Bernard J. Jansen](#)

Computer Science Program  
University of Maryland (Asian Division)  
Seoul, 140-022 Korea

## Abstract

This paper presents findings from a study of the effects of query structure on retrieval by Web search services. Fifteen queries were selected from the transaction log of a major Web search service in simple query form with no advanced operators (e.g., Boolean operators, phrase operators, etc.) and submitted to 5 major search engines - Alta Vista, Excite, FAST Search, Infoseek, and Northern Light. The results from these queries became the baseline data. The original 15 queries were then modified using the various search operators supported by each of the 5 search engines for a total of 210 queries. Each of these 210 queries was also submitted to the applicable search service. The results obtained were then compared to the baseline results. A total of 2,768 search results were returned by the set of all queries. In general, increasing the complexity of the queries had little effect on the results with a greater than 70% overlap in results, on average. Implications for the design of Web search services and directions for future research are discussed.

## Introduction

Information retrieval (IR) system searchers seldom use advanced searching techniques, such as Boolean operators or phrase searching ([Borgman, 1996](#); [Spink, et al., in press](#)). This has been especially true for Web searchers. The vast majority of Web searchers make little use of advanced query techniques. [Keily](#) (1997) conducted a study utilizing queries from two Web search engines, [WebCrawler](#) and [Magellan](#). Of the 2,000 queries, only 12% contained Boolean operators. [Hoelscher](#) (1998) presented data and analysis from [Fireball](#), a German Web IR system. The data set contained approximately 16 million queries. Three percent (3%) of the queries contained Boolean operators, and 8% contained phrase searching. Approximately 25% contained the 'must appear' operator, which was the plus sign (+). [Silverstein, et al.](#), (1999) presented results from an analysis of just under one billion queries submitted to the [Alta Vista](#) search engine. The researchers do not report the specific occurrence of Boolean operators, but 20.4% of the queries contained some advanced query operator (i.e., +, -, &, etc.). [Jansen, et al.](#) (2000) published a study concerning searching on the [Excite](#) search engine. In this analysis, approximately 8.5% of the queries contained Boolean operators. Approximately 9% of the queries contained some other advanced query operator.

[Jones, et al.](#) (1998) published research that focused on the [New Zealand Digital Library](#), a collection of technical computer science documents. They reported Boolean occurrence in over 25% of the queries, which is a substantially higher usage than reported in other studies on Web searching. This figure may be the result of the technical nature of the Web site's document collection.

The use of Boolean operators in these Web searching studies is substantially lower than the rates reported in studies of searchers using traditional information retrieval (IR) systems such as DIALOG or LEXUS/NEXUS. For example, [Siegfried, et al.](#) (1993) reported Boolean usage of over 36% on the DIALOG system, which the researchers considered a low rate of usage.

The implication of these findings leads to the question: Do Web searchers increase the probability of finding

relevant information by increasing the complexity their queries? Advanced searching techniques are well known and one can find numerous articles on advanced searching strategies ([Dragutsky, 1998](#)), tutorials on searching training ([Sullivan, 2000](#)) and educational classes on searching strategies ([UC Berkeley, 1997](#)). However, based on the Web studies referenced above, it appears that the majority of Web searchers continue to use very simple queries.

Recent studies suggest that Web users are finding the information they want. A survey of users on a major Web search engine reports that almost 70% of the users 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. Of the top ten Web sites in June 2000, 8 were Web search engines ([CyberAtlas, 2000](#)), implying that search engines are at least the best alternative available for finding information on the Web. Obviously, something is amiss. Searchers appear to be finding information using a technique that should be ineffective or at least inefficient. To shed some light on this apparent paradox, this study investigates the effect of complex queries (i.e., those using advanced syntax, such as Boolean operators) on the results retrieved by Web search services relative to the results retrieved by simple queries (i.e., those with no advanced syntax).

## Research Design

### Selection of Queries

Key to this research was the selection of queries that accurately reflected the structure of Web queries. Research shows that Web queries generally have two terms ([Jansen, et al., 2000](#); [Silverstein, et al., 1999](#)), cover a variety of topics ([Wolfram, 1999](#)), and are primarily noun phrases ([Jansen, et al., 2000](#); [Kirsch, 1999](#)). These criteria were used in selecting queries from an Excite search service transaction log. The 54,573 queries in the transaction log were collected on 10 March 1997. The specific query lengths reported by [Jansen, et al.](#) (2000) are listed in Table 1. The mean for query length (i.e., the number of terms in the query) was 2.21 terms with a standard deviation of 1.05 terms. These statistics are in line with those reported by other Web studies ([Silverstein, et al, 1999](#)) and presentations on Web searching data ([Kirsch, 1999](#); [Xu, 1999](#)).

Terms in query	Number of queries	Percent of all queries
More than 6	1,018	2
6	617	1
5	2,158	4
4	3,789	7
3	9,242	18
2	16,191	32
1	15,854	31
0	2,584	5
<b>Total</b>	51,433	100

Table 1: Query Length

Based on the information in Table 1, approximately 93% of the Web queries contained between 0 and 4 terms, inclusive. Since it is not meaningful to add query operators to queries of 0 or 1 terms, this study focused on queries with lengths between 2 and 4 terms inclusive. This range represents a majority of Web queries, or approximately 57%. Dropping queries containing 5 or more terms is also justifiable. First, only a small number of queries (7%) fall into this category. Second, longer queries are most likely constructed by a subset of Web searchers using more

sophisticated search techniques and would better be addressed in a separate study.

Based on the general distribution from Table 1, queries of the following lengths were selected for this study: 1 query of 4 terms, 3 queries of 3 terms, and 11 queries of two terms. Queries were selected on a variety of topics, since it has been reported that some search engines specialize in certain areas ([Neilsen/NetRating, 2000](#)). All queries that appeared on the popular query lists ([Searchwords, 2000](#)) or that referenced popular entertainers, popular locations, popular songs, etc. were eliminated, since Web search engines sometimes cache results from these highly queried topics ([Lesk, et al, 1997](#)). For similar reasons, all queries that were obviously queries for pornography were eliminated. With these goals and constraints as guides, the 15 queries displayed in Table 2 were selected.

Number of Terms	Query
4	nicotine levels smokeless tobacco
3	attention deficit disorder
3	flood plains definitions
3	ice cream cones
2	bikini thong
2	christmas scenes
2	dog crate
2	physical therapist
2	rhubarb pie
2	school buses
2	search engines
2	social workers
2	trumpet winsock
2	welfare state
2	time travel

Table 2: Queries from Excite Transaction Log.

Studies show that most Web searchers never view more than 10 results. [Hoelscher](#) (1998) reports that approximately 59% of the searchers viewed no more than 10 results. [Jansen, et al.](#) (2000) found that over 58% of the searchers viewed 10 documents or less. Similarly, [Silverstein et al.](#) (1999) reported that approximately 85% of the Web searchers viewed no more than 10 results. Based on this overwhelming evidence of Web searcher behavior, only the first 10 results in the results list were selected for comparison. Relevance judgements were not made concerning the results. The ability of Web search engines to retrieve relevant documents has been investigated several times ([Leighton & Srivastava, 1999](#)). In terms of quality, [Zumalt and Pasicznyuk](#) (1998) show that the utility of the Web may match that of a professional reference librarian.

## Searching Environment

Many Web sites offer searching features; however, this research focuses specifically on Web search engines. Search engines are the major portals for users of the Web, with 71% of Web users accessing search engines to locate other Web sites ([CommerceNet/Nielsen Media, 1997](#)). One in every 28 (3.5%) page views on the Web is a search results page ([Alexa Insider, 2000](#)). Search engines are, without a doubt, the IR systems of the Web. There are approximately 3,200 search engines on the Web ([Search Engine Watch, 2000](#)). However, only a handful of these dominate the market in terms of Web traffic. Among the better known, and those utilized in this research are Alta Vista, Excite, FAST Search, InfoSeek, and Northern Light.

## Document Collections

To understand searching on the Web, it is important to have a clear understanding of the size of the document collections involved. These Web search engines have individually indexed document collections that number in the millions of pages. To facilitate comparison, the document collection sizes are graphically displayed in the Figure 1.

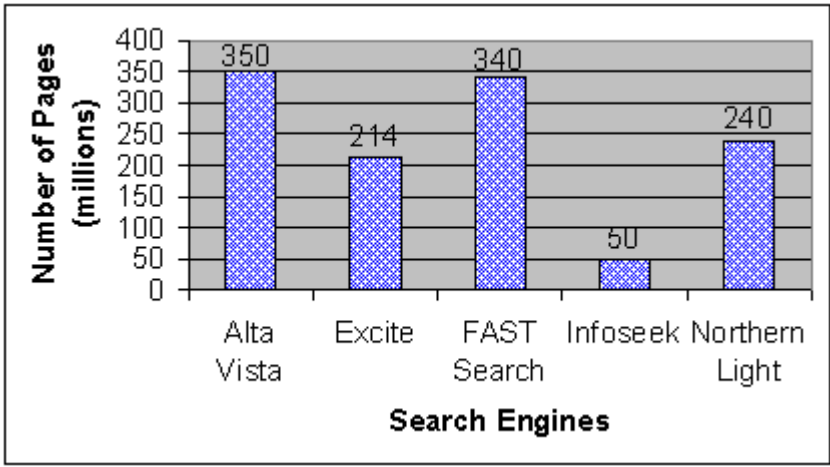


Figure 1: Size of Search Engine Document Collections (millions of pages)

Even with this magnitude, none of these document collections indexes the entire Web. Current estimates on the size of the Web range from approximately 350 million active Web pages ([Nielsen/Net Rating, 2000](#)), to 500 million non-duplicate Web pages ([FAST Search, 2000](#)) to approximately 800 million Web pages ([Lawrence & Giles, 1999](#)). So, it is difficult to say precisely what percentage of the Web these search engines cover. Regardless of the exact number, it is clear they all have indexed several million pages. Referring to Figure 1, Alta Vista and FAST Search have the largest collections at approximately 340 million pages. Excite and Northern Light have document collections of 214 and 240 million pages, respectively. Infoseek trails the pack with approximately 50 million pages. The sheer size of these collections has a unique impact on Web searching, where searchers can easily be inundated with results. Therefore, the focus of searching on the Web is primarily a precision-based service ([Xu, 1999](#)).

## Number of Visitors

In addition to the size of the document collections indexed, the reach of these search engines in number of searchers is also an important characteristic of Web searching. The number of unique visitors to these search engines varies widely although most attract a high number of visitors, as illustrated in Figure 2.

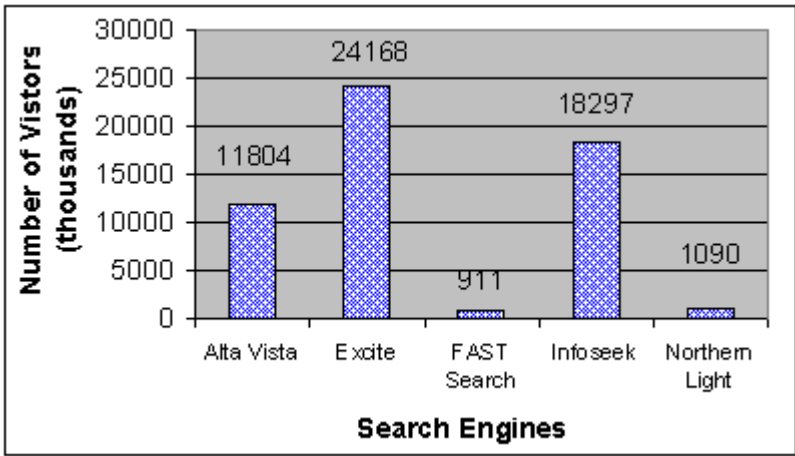


Figure 2: Number of Unique Visitors Per Month (thousands).

Notes: The number of monthly unique visitors to Northern Light and FAST Search is estimated based on audience reach data from [Alexa Research](#) (2000).

The data in Figure 2 was collected from [Alexa Research](#) (2000), [Nielsen/NetRating](#) (2000), and [CyberAtlas](#) (2000)

and represents the unique visitors to each site as of April 2000. The number of unique visitors is only a rough proxy of number of queries because not all visitors may submit a query and some visitors may submit multiple queries. See [Sullivan](#) (2000) for a discussion on the difficulties of estimating the actual number of searches per site. However, if one assumes that the percentage of actual searches relative to the total number of visits to a Web search engine is relatively constant across all search engines, one can get a general idea of the popularity of a search engine compared to the others. With thousands of unique visitors per months, these search services must be able to respond to a wide variety of topics and information needs.

## Searching Rules

The five search engines offer a variety of advanced searching options. For this research, only those advanced searching options that were available from the search engine's main page were utilized. Of the five search engines used, 4 offered all search options from the main page. One search engine, Alta Vista, does not support Boolean operators on the main page. In the case of Alta Vista, those search options were not investigated.

## Research Structure

All simple queries and appropriate complex queries were submitted to the five search engines on 21 May 2000. The terms from the original queries were all lower case. There were 75 simple queries and all returned at least 10 results. As mentioned earlier, results beyond the first 10 were discarded, providing 750 results to use as the baseline. Each of the 15 simple queries was then modified with the advanced searching operators supported by the various search engines. Many search engines offer drop down boxes (e.g., language of results, document collections to search) for refining the search. When drop down boxes were present on the main search page, the default options were utilized. In addition, "Power Searching" screens offered by some search engines were not utilized. Instead, all advanced queries were submitted via the search engine's main search screen.

## Results

A total of 150 complex queries was submitted. Each search engine offers different search options; therefore, the number of complex queries varied for each search engine, as outlined in Table 3.

			Search Options Supported			
Search Engine	Number of Simple Queries	Number of Complex Queries	+	"	AND	OR
Alta Vista	15	30	X	X		
Excite	15	60	X	X	X	X
FAST Search	15	30	X	X		
Infoseek	15	30	X	X		
Northern Light	15	60	X	X	X	X
<b>Total Queries</b>	75	210	75	75	30	30

Table 3: Number of Complex Queries by Search Engine.

From Table 3, all search engines supported the must appear (+) and phrase searching (") operators for a total of 75 queries each. Excite and Northern Light support Boolean operators AND and OR for a total of 30 queries each. Of the 210 complex queries, 201 returned 10 or more results. There were 9 queries that returned fewer than 10 results, all of which were phrase searching. One query returned 1 result, and one query returned 7 results. The remaining 7 queries returned 0 results and were not used in the comparison analysis. All told, there were 2,018 results returned by the complex queries. Combined with the 750 results from the simple queries, this gives 2,768 results.

In comparing the results between the simple and complex queries, the match had to be exact. The documents listed had to be the identical page at the same site. Different pages from the same site were not counted as matches. If a result appeared in both lists but in a different order, they were counted as a match as long as both were displayed in the top ten.

## Simple Versus Complex Query Comparison

The aggregate results of the analysis of the 2,768 results are display in Table 4.

Category	Average Number of Results that Appear in Baseline	Standard Deviation	Mode
Simple Queries	10.0	0.0	10
Complex Queries	7.3	1.3	10

Table 4: Comparison of Simple versus Complex Queries

Reviewing the statistics in Table 4, the baseline mean for the simple queries was 10, and the mean for the complex queries was 7.3. This means that, on average, 7.3 of the 10 results retrieved by the complex queries also appeared in the baseline results for the corresponding simple query on that search engine.

## Results by Search Engine

The comparison was also conducted for each search engine. These results are displayed in Table 5.

Search Engine	Average Number of Results that Appear in Baseline	Standard Deviation	Mode
Alta Vista	6	4.2	10
Excite	9	2.5	10
FAST Search	8	2.7	10
Infoseek	7	3.2	9
Northern Light	6	3.4	10

Table 5: Comparison of Results by Search Engine.

From examining Table 5, we see that Excite, FAST Search, and Infoseek will on average return 7 to 9 results exactly the same regardless of whether the query is simple or complex. Alta Vista and Northern Light means are slightly lower at 6. Alta Vista and Northern Light also have the largest standard deviations at 4.2 and 3.4 respectively.

## Results by Query Operator

The comparison was also conducted for each search operator. These results are displayed in Table 6.

Query Operator	Average Number of Results that Appear in Baseline	Standard Deviation	Mode
+	7.3	3.6	10
"	7.8	2.7	10
AND	6.8	3.7	10

<b>OR</b>	7.2	3.6	10
-----------	-----	-----	----

Table 6: Comparison of Results by Query Operator.

Table 6 shows that the highest correlation with the baseline results is with the phrase searching operator, the quotation mark ("). With this operator, almost 8 of the 10 results would also appear in the results list without the advanced search operator. With the must appear operator (+), Boolean intersection operator (AND) and Boolean union operator (OR), approximately 7 of the documents in the results list would have appeared without the use of the advanced operators.

## Results by Query

The analysis was also conducted for each query. These results are displayed in Table 7.

<b>Query</b>	<b>Average Number of Results that Appear in Baseline</b>	<b>Standard Deviation</b>	<b>Mode</b>
rhubarb pie	8.6	2.4	10
search engines	8.6	2.7	10
trumpet winsock	8.5	2.2	10
social workers	8.3	2.8	10
attention deficit disorder	8.1	2.2	10
physical therapist	8.1	2.5	10
welfare state	7.8	4.2	10
school buses	7.6	3.2	10
christmas scenes	7.4	3.3	10
bikini thong	6.9	3.5	10
dog crate	6.9	3.3	10
ice cream cones	6.6	3.6	10
time travel	6.6	4.1	10
nicotine levels smokeless tobacco	5.3	3.7	3
flood plains definitions	3.9	4.4	0

Table 7: Comparison of Results by Query.

The highest occurrence of overlap between the simple and complex results lists occurred with the queries *rhubarb pie*, *search engines*, *trumpet winsock*, *social workers*, *attention deficit disorder* and *physical therapist*. On average, about 8 of the 10 results for these queries were identical regardless of the presence or absence of advanced query syntax. At the other end of the spectrum, there was an overlap of approximately 4 results between the simple and complex queries with the query *flood plains definitions*. It is interesting to note that the two queries with substantially lower modes had more than two terms.

## Discussion

Referring to the data in Table 4, a paired t-test shows that the results from the simple queries are statistically significantly different from the results for complex queries. However, as with all statistics, one must ask what different does this make in the 'real world?'

Does it make sense to learn and utilize the more complex searching operators if on average it is only going to present the searcher with 2.7 different results than retrieved by just typing in the query? Are the 2.7 new results worth the increased chance of entering a query incorrectly? As the complexity of queries increase so does the probability of error ([Jansen & Pooch, in press](#)).

Findings suggest that the use of complex queries is not worth the trouble. Based on their conduct, it appears that most Web searchers do not think it is worth the trouble either. The behavior of Web searchers adheres to the



principle of least effort ([Zipf, 1948](#)), which postulates that there are "useful" behaviors that are quick and easy to perform. The very existence of these quick, easy behavior patterns then cause individuals to choose them, even when they are not necessarily the best behavior from a functional point of view. However, they are good enough, and people will generally expend the least amount of effort to achieve what they want. This can explain the behavior of Web searchers. The results obtained via simple queries are good enough.

The use of simple queries versus complex queries is also compelling when one compares the modes. The modes for the simple and the complex queries are both 10, meaning that more than any other occurrence, the results from a simple and complex queries will be the same. In reviewing the analysis by search engine, there was a great deal of overlap between query results for all search engines, ranging from 60% for Alta Vista and Northern Light to 90% for Excite. The mode for each of the search engines was 10, with the exception of Go/Infoseek with a mode of 9. With results being similar up to 90% of the time (e.g., Excite), one wonders why even have advanced searching syntax at all? Studies and presentations show that the failure rates among Web searchers using advanced syntax is high ([Jansen, et al., 2000](#); [Xu, 1999](#)). Why give searchers the opportunity to make mistakes? This seems to be the tactic followed by FAST Search and Go / Infoseek that limit the searcher's options.

In the analysis of the various advanced search operators, all had means of about 7, meaning on average approximately 7 of the 10 results were the same regardless of how simple or complex the query. The mode for all operators was 10. It appears that no particular operator has a significantly greater or lessor impact on results. The mean for phrase searching was a little higher at 7.8 and would have been a little higher except for a couple of the queries. For example, one can make a good case that *nicotine levels smokeless tobacco* is not a typical phrase that one would search for. However it is difficult to judge what searchers will do. It is also interesting that the results from the must appear operator (+) and the Boolean intersection operator (AND) are not the same. With Excite, the results were identical. With Northern Light, the results varied between the two operators.

In comparing individual queries, 6 had means of over 8 results, meaning that regardless of what advanced syntax was used, at least 8 of the results were the same as the baseline on average. Of these six queries, 1 was a three-term query and 5 were two-term queries. The query *flood plains definitions* has a mean of 3.9, substantially lower than all other queries. It was the only query with a mean less than 5. Even though it was a three-term query, the means of the other three-term queries were much higher at 8.1 and 6.6. Perhaps the topic or term choice was the determinate factor.

The four-term query was also near the bottom of the list with a mean 5.3 and a mode of 3. Still, more than 50% of the results were the same regardless of how the query was entered. With the four-term and one of the three-term queries at the bottom of the list, it may indicate that increased query length will increase the probability of different results from more complex queries. However, increasing the query length would perhaps also mean a review of what advanced syntax operators are appropriate. For example, as the query length increases phrase searching may no longer be a viable operator. In fact, if we remove the effect of phrase searching from these two queries, the mean for *nicotine levels smokeless tobacco* increases to 5.8 and the mean for *flood plains definitions* increases to 4.6.

## Conclusions and Future Research

This research indicates that the typical Web searcher is adhering to a very reasonable course of action by entering simple queries. The use of more complex queries appears to have a very small impact on the results retrieved. On average, 7.3 of the top ten results will be the same, regardless of how the query is entered. The two or three different results may not be worth the increased effort required to learn the advanced searching rules or the increased risk of making a mistake.

These results imply that Web search engine designers are doing a proper job of designing Web interfaces and ranking algorithms that accommodate the searching patterns of their customers. Some have criticized the Boolean model as being too complex for most users ([Salton, et al., 1993](#)). On most Web search engines, the current interface consists of a search text box and a search button. It is difficult to conceive of a simpler interface given current software and hardware technology. The results also call into question the strategy by some Web search services of having separate search pages for the Boolean and proximity operators. By forcing the more sophisticated users to go to another window, it may be discouraging the use of these advanced operators. Of course, this may be the goal.

This research also supports reviews that implementations of Boolean searching have many positive features that



overcome the shortcoming of the Boolean model. These practical features are sometimes ignored in the theoretical criticism of Boolean systems (Frants, *et al.*, 1999). Additionally, it validates the position that the shortcomings of Boolean systems, while theoretically valid, have limited practical impact (Korfhage, 1997), given the manner in which most people search.

The ranking algorithms of the Web search services are also supportive of the typical usage pattern of Web searchers. Based on the results of this research, one can conjecture that the ranking algorithms of these search engines adhere to the following rule: *Place at the top of the results list, those documents that contain all the query terms and that have all the query terms near each other.* This seems to be a reasonable course of action. With a ranking rule like this, the use of advanced query syntax will have little impact on the results at the top of the list.

There are several avenues for future investigation. The first is the use of a larger set of initial queries, increasing the number of simple queries from 15 to say 100. A larger set of simple queries would increase the diversity of search subjects and query terms. The second would be to examine the effect of longer queries. This research focused on the dominant behavior of Web searchers, short queries (generally about 2 terms) and viewing 10 or fewer documents. It would be interesting to see if these same results hold as the query length or the number of results viewed increase. However, the impact in terms of explaining Web searching behavior would be less. Finally, an exciting research area to explore would be the use negative operators and ways to present them to the searcher. Some search engines now offer online thesauruses that automatically suggest terms for the user to add to the query. It would be interesting to also offer term suggestions that the searcher may not want in the results. These terms could then be added to the query using the must not appear operator, usually a minus sign (-) or the Boolean operators (AND NOT).

## References

- Alexa Insider (2000). Alexa insider's page. Available at: <http://insider.alexa.com/insider?cli=10>. [30 March 2000]
- Alexa Research. (2000). [The Alexa 1000](http://www.alexaresearch.com/clientdir/web_reports/top_websites_nonclient.cfm). Available at: [http://www.alexaresearch.com/clientdir/web\\_reports/top\\_websites\\_nonclient.cfm](http://www.alexaresearch.com/clientdir/web_reports/top_websites_nonclient.cfm) [30 August 2000]
- Borgman, C. (1996). Why are online catalogs still hard to use? *Journal of the American Society for Information Science*. **47**(7), 493-503.
- CommerceNet/Nielsen Media. (1997). Search engines most popular method of surfing the Web. Available at: <http://www.commerce.net/news/press/0416.html> [30 August 2000]
- CyberAtlas. (2000) Available at: <http://www.cyberatlas.com> [30 May 2000]
- Dragutsky, P. (1997). *Boolean basics*. Available at: [http://www.suite101.com/article.cfm/search\\_engines/4825](http://www.suite101.com/article.cfm/search_engines/4825) [30 August 2000] (Note: you need to scroll down the page to find this item; there is no internal link.)
- FAST Search (2000). *FAST Search FAQ*. Available at: <http://www.ussc.alltheweb.com/faq.php3> [1 May 1999]
- Frants, V., Schairo, J., Taksa, I. & Voiskunskii, V. (1999). Boolean search: current state and perspectives. *Journal of the American Society for Information Science*. **50**(1): 86-95.
- Hoelscher, C. (1998). How Internet experts search for information on the Web. *World Conference of the World Wide Web, Internet, and Intranet, Orlando, FL., 1998*. Association for the Advancement of Computing in Education. (Note: Published on CD-ROM with no page numbers.)
- Jansen, B., & Pooch, U. (In Press). Web user studies: a review of and framework for future research. *Journal of the American Society for Information Science*. Forthcoming. Available in draft form at: <http://jimjansen.tripod.com/academic/pubs/wus.html>. [30th August 2000]
- Jansen, B. Spink, A., Bateman, J., & Saracevic, T. (1998). *Real life information retrieval: a study of user queries on the Web*. *SIGIR Forum*, **32**, 1, 5 - 17. Available at: <http://jimjansen.tripod.com/academic/pubs/sigirforum98/forum98.html>. [30th August 2000]
- Jansen, B., 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 Management*. **36**(2), 207-227.
- Jones, S., Cunningham, S. & McNab, R. (1998). Usage analysis of a digital library. *Third ACM Conference on Digital libraries, Pittsburgh, PA. 1998*. ACM SIGIR. 293 - 294. Available at <http://www.acm.org/pubs/citations/proceedings/dl/276675/p293-jones/>. (Note: You must be a subscriber of the ACM Digital Library to access the article.)
- Keily, L. (1997). Improving resource discovery on the Internet: the user perspective. *21th International Online Information Meeting, London, UK*, Learned Information Europe Ltd. 205 - 212.

- Kirsch, S. (1998). The future of Internet search (keynote address). Paper presented at the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Melbourne, Australia. Available at: <http://www.skirsch.com/stk.html/presentations/sigir.ppt> [16 August 1999]
- Korfhage, R. (1997). *Information Storage and Retrieval*. New York, NY: Wiley.
- Lawrence, S. & Giles, C. (1999). Accessibility of information on the web. *Nature*, **400**, 107 – 109.
- Leighton, H. and Srivastava, J. (1999). First 20 Precision among World Wide Web search services (search engines), *Journal of the American Society for Information Science* **50**(1), 870-881.
- Lesk, M., Cutting, D., Pedersen, J., Noreault, T., & Koll, M. (1997). Panel Session on "real world" information retrieval. *SIGIR Forum*, **32**, 1, 1 - 4.
- Nielsen//NetRatings (2000). Available at: <http://www.nielsen-netratings.com/> [11 August 1999]
- Search Engine Watch. (2000). Available at: <http://searchenginewatch.com> [30 August 2000]
- SearchWords.com (2000). Available at: <http://www.searchwords.com> [30 May 2000]
- Siegfried, S., Bates, M., & Wilde, D. (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, H. & Moricz, M. (1999). Analysis of a very large Web search engine query log. *SIGIR Forum*, **33**(1), 6 -12.
- Spink, A., Bateman, J., & Jansen, B. (1998). [Searching heterogeneous collections on the Web: behaviour of Excite users.](#) *Information Research*, **4**(2) Available at: <http://www.shef.ac.uk/~is/publications/infres/paper53.html> [30 August 2000]
- Spink, A., Wolfram, D., Jansen, B., & Saracevic, T. (In Press). The public and their queries. *Journal of the American Society for Information Science*. Forthcoming.
- Sullivan, D. (2000). Search engine sizes. Available at: <http://searchenginewatch.com/reports/sizes.html> [30 August 2000]
- University of California, Berkeley. Extension. (1997). [Online searching and electronic research. Lesson 6: Introduction to Boolean logic.](#) Berkeley, CA: University of California, Berkeley. Available at: [http://www.exo.net/uce/uce6\\_boolean.html](http://www.exo.net/uce/uce6_boolean.html). [30 August 2000]
- Wolfram, D. (1999). Term co-occurrence in Internet search engine queries: An analysis of the Excite data set. *Canadian Journal of Information and Library Science*. **24**(2/3), 12-33.
- Xu, J. L. (1999) Internet search engines: real world IR issues and challenges. Paper delivered at the *Conference on Information and Knowledge Management, Kansas City, Missouri*.
- Young, D. & Shneiderman, B. (1993). A graphical filter/flow model for Boolean queries. *Journal of the American Society for Information Science*, **44**(1), 327-339.
- Zumalt, J. & Pasicznyuk, R. (1998). The Internet and reference services: a real-world test of Internet utility. *Reference and User Services Quarterly*, **38**(2), 165 – 172.

---

**[Information Research, Volume 6 No. 1 October 2000](#)**

*The effect of query complexity on Web searching results*, by [Bernard J. Jansen](#) Location:

<http://www.shef.ac.uk/~is/publications/infres/paper87.html> © the author, 2000.

Last updated: 30th August 2000

---