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

A Study of Web Search Trends

Amanda Spink
School of Information Sciences, University of Pittsburgh, 610 IS Building, 135 N. Bellefield Avenue, Pittsburgh PA 15260, Tel: (412) 624-9454 Fax: (412) 648-7000

Bernard J. Jansen
School of Information Sciences and Technology, The Pennsylvania State University, 329F IST Building, University Park, PA 16802, Tel: (814) 865-6459 Fax: (814) 865-6426

Abstract

This article provides an overview of recent research conducted from 1997 to 2003 that explored how people search the Web. The article reports selected findings from many research studies conducted by the co-authors of the paper from 1997 to 2003 using large-scale Web query transaction logs provided by commercial Web companies, including Excite, Alta Vista, Ask Jeeves, and AlltheWeb.com. The many studies are also synthesized in the recent book "Web Search: Public Searching of the Web" by Amanda Spink and Bernard J. Jansen (Kluwer Academic Publishers). The researchers examined the topics of Web searches; how users search the Web using terms in queries during search sessions; and the diverse types of searches, including medical, sex, e-commerce, multimedia, etc. information. Key findings include changes in search topics since 1997, including a shift from entertainment to e-commerce queries. Further findings show little change in many aspects of Web searching from 1997-2003, including query and search session length. The studies also show more complex Web search behaviors by a minority of users who conduct multitasking and successive searches.

Keywords

Web searching, Web information retrieval, Web search behavior, Search engines

Introduction

People are spending increasing their electronic information access via Web search engines. Web searching services such as Google and Yahoo are now tools that people access everyday to find information. Studies that investigate such issues as Web searching trends are important for both users and Web search engines alike. User's Web search context can be examined at many levels, including the information environment/social level, organizational level, information seeking level, human-computer interaction level and query level. In order to understand to design better Web search tools, we need to understand more about how people use the Web at various levels.

For many people, Web interactions are often short but frequent. Many studies have explored the effectiveness of Web search engines (Lawrence & Giles, 1998) and begun to model how users search the Web (Silverstein, Henzinger, Marais, & Moricz, 1999; Wolfram, Spink, Jansen, & Saracevic, 2001). Web search engines are trying to better support human information behaviors through the development of a new Web search tools, to help users conduct better electronic information seeking.

This paper provides selected results from a large and ongoing research project exploring Web search behavior. These studies were also synthesized in the recent book "*Web Search: Public Searching of the Web*" by Amanda Spink and Bernard J. Jansen (Kluwer Academic Publishers). Our research focuses at the human-computer interaction and query level of Web search behavior. We report results from studies of Web query data from the Excite, AlltheWeb.com, Alta Vista and Ask Jeeves from 1997 to 2003. The researchers could not obtain Web query data from Google, Yahoo or MSN. Most recently Web query data was obtained from Alta Vista.com before it became part of Yahoo.com (Jansen, Spink & Pedersen, in press).

Our research goal has been to track trends in public Web searching and examine how the public searches the Web (Jansen, Spink & Saracevic, 2000; Spink, Wolfram, Jansen, & Saracevic, 2001).

Data Collection and Analysis

We analyzed many large sets of Web query data provided by various Web companies from 1997 to 2003. All Web search engine users were anonymous and could not be identified in any way. However, we could identify each user's sequence of queries. The transaction records analyzed contained three fields. With these three fields, we were able to locate a user's first query and recreate the chronological series of actions by each user in a session:

- *Time of Day*: measured in hours, minutes, and seconds;
- *User Identification*: an anonymous user code assigned by the Web server;
- *Query Terms*: exactly as entered by the given user.

Our analysis focused on three levels of data analysis--sessions, queries, and terms. This large-scale study of Web searching provides insights into Web searching.

Selected findings are provided below that provide interesting insights into patterns of public Web searching, including the structure of Web searches, what topics people search for, and search behavior in areas such as multimedia, sex and e-commerce.

SELECTED FINDINGS

Web Search Characteristics

Web Queries

Our analysis of Excite data shows that the mean length of Excite queries increased steadily from 1.5 in 1996 to 2.6 in 1999 and 2.4 by 2003, and the mean number of terms in unique queries was 2.4. The mean query length for U.S/U.K. users in 1996 was 1.5 and mean query length for European users in 1997 was 1.5--for 1999 U.S/U.K. users mean query length was 2.6, and for European users it was 1.9. English language queries increased in length more quickly than European language queries.

Jansen, Spink, and Saracevic (2000) report that Web queries were short and most users did not enter many queries per search. The mean number of queries per user was 2.8 in 1997.

Spink, Ozmutlu, Ozmutlu and Jansen (2002) also report some differences in US and European Web searching, including shorter US sessions.

However, a sizable percentage of users did go on to either modify their original query or view subsequent results. On the average, a query contained 2.21 terms in 1997. About one in three queries had one term only, two in three had one or two terms, and four in five had one, two, or three terms. Fewer than 4 percent of the queries were comprised of more than 6 terms. Spink, Jansen, Wolfram, and Saracevic (2002) reported the mean terms per query had increased slightly to 2.6 by 2001 and 2.4 by 2003. Overall, general Web queries are still short with most users entering 2-3 term per query and 2-3 queries per search.

Question and Request Format Web Queries

Spink and Ozmutlu (2002) report that some 50 percent of Ask Jeeves users submitted queries in question format. Most questions began with the words "Where do I find . . .? ". Some 25 percent of users phrased their queries in request format, most commonly "Get me information . . .". Overall, most general Web queries are in keyword query format rather than question format.

Search Terms: Distribution

Jansen, Spink, and Saracevic (2000) showed that the frequency distribution of the query terms is skewed. A few terms were used repeatedly and a lot of terms were used only once. On the top of the list, the sixty-three subject terms that had a frequency of appearance of 100 or more represented less than 1 percent of all terms, but they accounted for about 10% of all queries. Terms that appeared only once formed half of the unique terms.

By 2001, 615 of terms were not repeated in the dataset (Spink, Jansen, Wolfram & Saracevic, 2002). Our studies show that Web searching can be characterized by a small percentage of high-frequency terms and many low-frequency terms.

Web Query Reformulation

Spink, Jansen, and Ozmultu (2000) found that most Web search engine users searched one query only and did not follow with successive queries. The average session, ignoring identical queries, included about 1.6 queries. About two in three users submitted only one query, and six in seven did not go beyond two queries. Spink, Jansen, Wolfram, and Saracevic (2002) reported that in 2001 some 44 percent of users modified their queries with 25 percent of users entering three or more queries. Overall, most users still enter only one or two queries, and conduct limited query reformulation.

Use of Boolean Operators and Advanced Search Features

The use of Boolean operators (AND, OR, NOT, +, -) increased from 22 percent of queries in 1997 to 28 percent of queries in 1999. From the 1996-1999 data set, approximately 8 percent of searches included proximity searching. However, in another study Jansen, Spink, and Saracevic (2000) found that Boolean operators were seldom used. One in eighteen users used Boolean capabilities, and of the users employing them, every second user made a mistake, as defined by Excite rules.

However, a majority of these uses were mistakes (about two out of three). Spink, Jansen, Wolfram, and Saracevic (2002) reported that by 2001 some 10 percent of Web searches contained Boolean operators. Overall, we see that Boolean search is and the use of advanced search features is still limited.

Analysis of Web searches shows that when available relevance feedback, as an advanced search feature, is rarely used. In 1999 about one in twenty Excite queries used the feature "More Like This." Spink, Jansen, and Ozmultu (2000) found that one-third of Excite users went beyond the single query, with a smaller group using either query modification or relevance feedback, or viewing more than the first page of results.

Viewing Results

From 1996 to 1999, for more than 70 percent of the time, a user only viewed the top ten results. On average, users viewed 2.35 pages of results (where one page equals ten hits). Over 50% of the users did not access results beyond the first page. Jansen, Spink, and Saracevic (2000) found that more than three in four users did not go beyond viewing two pages. By 2001, only roughly one-third of users looked beyond the second page of Web sites retrieved (Spink, Jansen, Wolfram, & Saracevic, 2002). By 2003, in general users view about five Web documents per query (Jansen & Spink, 2003).

WEB SEARCH TOPICS

Types of Web Searches

People search the Web for an infinite variety of topics. We now examine how users search on particular topics such as sex, e-commerce, and medical information. Spink, Jansen, Wolfram, and Saracevic (2002) report a shift in Web search topics from entertainment and sex in 1997 to commerce, travel, employment, economy, people, places, and things in 2001. Search topics have shifted from entertainment to e-commerce as the content of the Web has shifted more towards business and people searching (Spink, Jansen & Pedersen, in press).

Sexually Related Searching

Jansen, Spink, and Saracevic (2000) found searching about sex on Excite represents only a small proportion of all searches but were the top frequency. One in every four terms in the highest used terms were sexual terms. However, sexual terms form a very small proportion of all terms. The diversity of subjects searched is very high. Spink, Ozmultu, and Lorence (2004) and Spink, Koricich, Jansen and Cole (2004) found that sexually related searches were longer than non-sexual searches and included viewing more pages of Web sites, particularly for images.

Medical and Health-related Web Searching

Spink et al. (2004) found that medical and health related searches form a small percentage of Web searching. The top five categories of medical or health advice sought were: general health, weight issues, reproductive health and puberty, pregnancy/obstetrics, and human relationships. From 1997 to 2003 Medical and health queries declined as a proportion of Web queries. One reason may be the increase in the use of specialized medical/health Web sites and the increase in e-commerce-related queries.

E-Commerce Searching

E-commerce queries have increased on Web search engines (Spink & Guner, 2001; Spink & Jansen, 2004). Spink and Guner (2001) found that business queries often include more search terms, are less modified, lead to fewer Web pages viewed, and include less advanced search features. Company or product name queries were the most common form of business query. Spink, Jansen, Wolfram, and Saracevic (2002) found that by 2001 the

largest category of Web searches were e-commerce-related. Jansen, Jansen and Spink (in press) examined job search Web queries.

Multimedia Searching

Goodrum and Spink (2001) conducted a specific analysis of image queries within the 1.2 million queries. Provisions for image searching by Web search engines are important for users. Users seeking images input relatively few terms to specify their image information needs on the Web. Users seeking images interact iteratively during the course of a single session but input relatively few queries overall. Most image terms are used infrequently with the top term occurring in less than 9 percent of queries.

Jansen, Spink, and Saracevic (2000) found that terms indicating sexual or adult content materials appear frequently in image queries. Overall, multimedia searching is shifting as the content of the Web changes (Jansen, Goodrum & Spink, 2000; Jansen, Spink & Pedersen, in press; Ozmutlu, Spink, & Ozmutlu, 2002).

MORE COMPLEX SEARCH PATTERNS

Despite the generally short nature or user Web queries and search sessions, recent studies are also showing that some users are engaging in more complex Web search interactions.

Multitasking Search

Spink, Ozmutlu, and Ozmutlu (2002) and Spink, Park and Jansen (in press) found that many Web searches involved users seeking information on two or more topics concurrently. Overall, we see some users moving towards more complex searches by a minority of users that involve multiple related interactions and multiple topics.

Successive Searching

Spink, Bateman, and Jansen (1999) conducted an interactive survey of over 300 Excite users and found that many had conducted two searches or three or more related searches using the Excite search engine over time when seeking information on a particular topic. Successive searches often involved a refinement or extension of the previous searches as new databases were searched and search terms changed as the Excite users' understanding and evaluation of results evolved over time from one successive search to the next.

Discussion

Our research from 1997 to 2003 shows some patterns and trends in general Web searching. In summary, most Web queries are short, without query reformulation or modification, and have a simple structure. Few search sessions include advanced search techniques, and when they are used, many include mistakes. Advanced search features are at best slowly growing in use. Many Web searches retrieve a large number of Web sites, but users' view few result pages and generally view about five Web documents. Overall, a small number of terms are used with high frequency and many terms are used only once in our data sets. Web queries are rich in subject diversity. Some Web search engine users are conducting more longitudinal Web searching during information seeking.

The Web is an interaction tool, yet many users Web interactions are limit, possibly due to a lack of user training. In general, Web search engines explain little to their users and do not tell users that Web search engines only cover a limited number of Web sites (Lawrence & Giles, 1998). Many Web search sessions are "quick and dirty" rather than an interactive

search process. Overall, the effectiveness of users' search interactions depends on more effective search techniques and greater user training.

Conclusion and Further Research

We continue to examine searches by Web search engine users using large-scale Web query transaction logs. Our studies continue to show some significant trends in general Web searching. Further research is needed to asses users' the performance of different Web search engines. However, our findings do provide a trend analysis of public Web searching that can help improve Web search engines.

Further research is currently being conducted using query data from other Web search engines, including examination of the use of advanced search features and their use of more complex search techniques. Further ongoing Web user behavior research is needed using data from different Web search engines, as well as studies of new approaches to user training, interfaces and software agents, and schemas to aid users to improve their Web searching.

References

- Goodrum, A., & Spink, A. (2001). Image searching on the Excite web search engine. *Information Processing and Management*, 37(2), 95-312.
- Jansen, B. J., Goodrum, A., & Spink, A. (2000). Searching for multimedia: An analysis of audio, video, and image Web queries. *World Wide Web: An International Journal*, 3(4), 249-254.
- Jansen, B. J., Jansen, K., & Spink, A. (2005). Using the Web to look for work: Implications for online job seeking and recruiting. *Internet Research: Electronic Networking Applications and Policy*, 15(1), 49-66.
- Jansen, B. J., & Spink, A. (2003). An analysis of web information seeking and use: Documents retrieved versus documents viewed. In *Proceedings of the 4th International Conference on Internet Computing*, pp. 65-69. Las Vegas, Nevada. 23-26 June.
- Jansen, B. J., & Spink, A. (2005). How are we searching the World Wide Web? A comparison of nine large search engine transaction logs. *Information Processing and Management*, 42(1), 248-263.
- Jansen, B. J., Spink, A., & Pedersen, J. (2005). The effect of specialized multimedia collections on Web searching. *Journal of Web Engineering*, 3(3/4), 182-199.
- Jansen, B. J., Spink, A., & Pedersen, J. (2005). A temporal comparison of AltaVista Web searching. *Journal of the American Society for Information Science and Technology*, 56(6), 559-570.
- Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users and real needs: A study and analysis of users' queries on the Web. *Information Processing and Management*, 36(2), 207-227.
- Lawrence, S., & Giles, C. L. (1998). Searching the World Wide Web. *Science*, 280(5360), 98-100.
- Ozmutlu, S., Spink, A., & Ozmutlu, H. C. (2003). Trends in multimedia web searching: 1997-2001. *Information Processing and Management*. 39(4), 611-621.
- Silverstein, C., Henzinger, M., Marais, H., & Moricz, M. (1999) Analysis of a very large Web search engine query log. *ACM SIGIR Forum*, 33, 3.
- Spink, A., Bateman, J., & Jansen, B. J. (1999). Searching the Web: Survey of EXCITE users. *Internet Research: Electronic Networking Applications and Policy*, 9(2), 117-128.
- Spink, A., & Guner, O. (2001, July). E-commerce Web queries: Excite and Ask Jeeves study. *First Monday*, 6(7).

- Spink, A., & Jansen, B. J. (2004). *Web Search: Public Searching of the Web*. Kluwer Academic Publishing.
- Spink, A., Jansen, B. J., & Ozmultu, H. C. (2000). Use of query reformulation and relevance feedback by Web users. *Internet Research: Electronic Networking Applications and Policy*, 10(4), 317-328.
- Spink, A., Jansen, B. J., & Pedersen, J. (2004). Searching for people on Web search engines. *Journal of Documentation*, 60(3), 266-278.
- Spink, A., Jansen, B. J., Wolfram, D., & Saracevic, T. (2002). From e-sex to e-commerce: Web search changes. *IEEE Computer*, 35(3), 133-135.
- Spink, A., Koricich, A., Jansen, B. J., & Cole, C. (2004). Sexual searching on Web search engines. *Cyberpsychology & Behavior*, 7(1), 65-72.
- Spink, A., & Ozmutlu, H. C. (2002). Characteristics of question format Web queries: An exploratory study. *Information Processing and Management*, 38(4), 453-471.
- Spink, A., Ozmutlu, H. C., & Lorence, D. P. (2004). Web searching for sexual information: An exploratory study. *Information Processing and Management*, 40(1), 113-124.
- Spink, A., Ozmutlu, H. C., & Ozmutlu, S. (2002). Multitasking information seeking and searching processes. *Journal of the American Society for Information Science and Technology*, 53(8), 639-652.
- Spink, A., Ozmutlu, S., Ozmutlu, H. C., & Jansen, J. (2002). US versus European Web searching trends. *ACM SIGIR Forum*, 36(2).
- Spink, A., Park, M., & Jansen, B. J. (2006). Multitasking during Web search sessions. *Information Processing and Management*, 42(1), 264-275.
- Spink, A., Wolfram, D., Jansen, B. J., & Saracevic, T. (2001). Searching the Web: The public and their queries. *Journal of the American Society for Information Science*, 53(2), 226-234.
- Spink, A., Yang, Y., Nykanen, P., Lorence, D. P., Jansen, B. J., Ozmutlu, S., & Ozmutlu, H. C. (2004). A study of medical and health queries to a Web search engine. *Health Information and Libraries Journal*, 21(1), 44-51.
- Wolfram, D., Spink, A., Jansen, B. J., & Saracevic, T. (2001). Vox populi: The public searching of the Web. *Journal of the American Society for Information Science and Technology*, 52(12), 1073-1074.

Bibliographic information of this paper for citing:

Spink, A., & Jansen, B.J. (2004). "A study of Web search trends". *Webology*, **1**(2), Article 4. Available at: http://www.webology.org/2004/v1n2/a4.html

---

This article has been cited by other articles.