Web Search: Emerging Patterns

AMANDA SPINK

ABSTRACT

THIS ARTICLE EXAMINES the public searching of the Web and provides an overview of recent research exploring what we know about how people search the Web. The article reports selected findings from studies conducted from 1997 to 2002 using large-scale Web user data provided by commercial Web companies, including Excite, Ask Jeeves, and AlltheWeb.com. We examined what topics people search for on the Web; how people search the Web using keywords in queries during search sessions; and the different types of searches conducted for multimedia, medical, e-commerce, sex, etc., information. Key findings include changes and differences in search topics over time, including a shift from entertainment to e-commerce searching by largely North American users. Findings show little change in current patterns of Web searching by many users from short queries and sessions. Alternatively, we see more complex searching behaviors by some users, including successive and multitasking searches.

INTRODUCTION

People are spending increasing amounts of time working with electronic information. Web searching services such as Alta Vista and Google are now everyday tools for information seeking.

The research that explores such issues as the organization of the Web or Web searching trends is becoming more important for users and Web search engines alike. There are many overlapping and related levels of a user's context that are relevant to Web research, including the information environment/social level, organizational level, information-seeking

level, human-computer interaction level, and query level. In order to better understand how to organize the Web, we also need to understand more about how people interact with and use the Web at these different levels.

For many users, Web interactions are often frustrating and con-strained. A growing body of large-scale quantitative or qualitative studies is exploring these issues, including the effectiveness and limitations of Web search engines (Lawrence & Giles, 1998) and how users search the Web (Silverstein et al., 1999; Wolframn et al., 2001). One outgrowth of Web research is better support for human information behaviors and the devel-opment of a new generation of Web tools, such as Web meta–search engines, to help users persist in electronic information seeking and help people resolve their information problems.

This article reports selected results from a large-scale and ongoing series of studies of searching behavior on commercial Web search engines by a diverse range of users. The research reported in this article is focused at the human-computer interaction and query level of Web user behavior. Selected results are reported from studies of Web query data from Excite, AlltheWeb.com, and Ask Jeeves. The researchers were not able to obtain data from the major Web company Google, but further analysis is being conducted on Web query data from Alta Vista. The goal of these studies is to track trends in the public searching of the Web and explore how the public searches the Web (Spink, Wolfram, Jansen, & Saracevic, 2001).

WEB QUERY DATA SETS

The analysis was conducted on various large sets of Web query data provided by various Web companies from 1997 to 2001. All users were anonymous and could not be identified in any way. But we could identify each user's sequence of queries.

Each transaction record contained three fields. With these three fields, researchers were able to locate a user's initial 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.

We focused on three levels of data analysis—*sessions, queries,* and *terms.* This large-scale study provides insights into Web searching with implica-tions for developing better search engines and services.

WEB SEARCH PATTERNS

Selected findings, summarized below, provide interesting insights into current patterns of public Web searching, including how people structure

their Web searches, what they search for, and search behavior in special topic areas.

Web Queries

How long are general Web queries? The mean length of Excite queries increased steadily from 1.5 words in 1996 to 2.6 in 1999, 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—in 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.

However, a sizable percentage of users did go on to either modify their original query or view subsequent results. On 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 six terms. Spink, Jansen, Wolfram, and Saracevic (2002) reported the mean terms per query had increased slightly to 2.6 by 2001. Overall, general Web queries are still short.

Use of Boolean Operators

How frequently are Boolean operators used during Web searching? 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 1996–99 data set, approximately 8 percent of searches included proximity searching. Jansen, Spink, and Saracevic (2000) found that Boolean operators were seldom used. One in eighteen users used any Boolean capabilities and, of the users employing them, every second user made a mistake, as defined by Excite rules. The '+' and '-' modifiers that specify the mandatory presence or absence of a term were used more than Boolean operators. About one in twelve users employed them. About one in eleven queries incorporated a '+' or '-' modifier. But 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 still in limited use.

Web Query Reformulation

Do Web search engine users reformulate their queries? Spink, Jansen, and Ozmultu (2000) found that most users searched one query only and did not follow with successive queries. The average session, ignoring identical queries, included 1.6 queries. About two in three users submitted a

single 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 little query reformulation.

Question and Request Format Web Queries

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

Search Terms: Distribution

What is the distribution of search terms? Jansen, Spink, and Saracevic (2000) report the distribution of the frequency of use of terms in queries as highly skewed. A few terms were used repeatedly and many 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 only one-third of 1 percent of all terms, but they accounted for about one of every ten terms used in all queries. Terms that appeared only once amounted to half of the unique terms. By 2001, 615 terms were not repeated in the dataset, as reported by Spink, Jansen, Wolfram, and Saracevic (2002). Overall, Web searching involves a small percentage of high-frequency terms and many low-frequency terms.

Use of Relevance Feedback

How frequently are relevance feedback commands used? Analysis of Web searches shows that, when available, relevance feedback is rarely used. About one in twenty 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. They examined the occurrence of each query type (unique, modified, relevance feedback, view a results page, etc.) in a large sample of user sessions. The distribution of query type changes as the length of the user session increases. For the user sessions of two and three queries, the relevance feedback query is dominant. As the length of the sessions increase, the occurrences of relevance feedback as a percentage of all query types decreases. Some 63 percent of relevance feedback sessions could be construed as being successful. If the partially successful user sessions are included, then more than 80 percent of the relevance feedback sessions provided some measure of success.

Viewing Results

How many pages of ten hits do users view? This is a very interesting question for users and Web industry people alike. 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 half 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).

WEB SEARCH TOPICS

Users search the Web on an infinite variety of topics. The next section focuses on what we know about 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 toward business.

Sexually Related Searching

Jansen, Spink, and Saracevic (2000) found searching about sex on Excite represents only a small proportion of all searches. When the top frequency terms are classified as to subject, the top category is "Sexual." As to the frequency of appearance, about one in every four terms in the list of sixty-three highest used terms can be classified as sexual in nature. But while sexual terms are high as a category, they still represent a very small proportion of all terms. Many other subjects are searched and the diversity of subjects searched is very high.

Spink, Ozmultu, and Lorence (in press) found that sexually related searches were longer than general searches and involved viewing more pages of Web sites. Overall, sexual Web searchers are more persistent and likely to be seeking images.

Medical and Health-related Web Searching

Medical and health-related information is proliferating on the Web. Spink, Yang, Nykanen, Lorence, Ozmutlu, and Ozmutlu (in press) found that a small percentage of Web searching is medical or health-related. The top five categories of medical or health advice sought were general health, weight issues, reproductive health and puberty, pregnancy/obstetrics, and human relationships. Trends show that medical and health queries have declined as a proportion of Web queries as the use of specialized medical/health Web sites and e-commerce–related queries has increased, but e-commerce–related searching has increased substantially.

E-Commerce Searching

E-commerce queries are increasing on the Web (Spink & Guner, 2001). Web queries are a primary means for translating people's business product, service, and information needs for e-commerce. Spink and Guner (2001) found that business queries often include more search terms than other types of queries, 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. The most common business-related query submitted to Ask Jeeves was "Where can I buy . . . " or the request "I want to buy . . . " Spink, Jansen, Wolfram, and Saracevic (2002) found that by 2001 the largest category of Web searches were e-commerce related.

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 many terms were unique in the large data sets, with over half of the terms used only once. Terms indicating sexual or adult content materials appear frequently in image queries. They represented a quarter of the most frequently occurring terms but were a small percentage of the total terms. Overall, multimedia searching is shifting as the content of the Web changes (Jansen, Goodrum, & Spink, 2000; Ozmutlu, Spink, & Ozmutlu, 2002).

LONGITUDINAL SEARCH PATTERNS

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

Successive Searching

How many Web searches do users conduct on a particular topic? Spink, Bateman, and Jansen (1999) conducted an interactive survey of over three hundred 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.

Multitasking Search
    How many topics are users searching for? Spink, Ozmutlu, and
Ozmutlu (2002) found that many Web searches involved users seeking
information on two or more topics concurrently. Overall, we see some
users moving toward more complex searches that involve multiple related
interactions and multiple topics.

DISCUSSION
    The research we conducted over the last five years shows some inter-
esting patterns and trends in general Web searching. In summary, most
Web queries are short, without much modification, and simple in struc-
ture. Few queries incorporate advanced search techniques and, when they
are used, many mistakes result. However, advanced search features are
slowly growing in use. Many people retrieve a large number of Web sites,
but view few results pages and tend not to browse beyond the first or sec-
ond results pages. Overall, a small number of terms are used with high fre-
quency and many terms are used once. Web queries are very rich in subject
diversity, and some are unique. The subject distribution of Web queries
does not seem to map to the distribution of Web sites' subject content.
Some users are engaging in more longitudinal Web searching practices
during their information-seeking processes that are not well supported by
Web search technologies. We can see that Web searching is growing as a
huge public challenge, but it is an imprecise and challenging skill.
    Insights into Web searching trends and patterns have implications for
the organization of the Web. A key problem for Web organization is that
people in general do not really understand how Web search engines work
or the structure of the Web. The Web is a creature of interaction, yet many
Web interactions are subject to limitations due to a lack of information and
training by users. In general, Web search engines do not explain the Web
to users and do not tell users that their search engines only cover a limited
number of Web sites. Web culture is based on a "quick and dirty" approach
to searching, rather than an exploratory, interactive approach. Web orga-
nizational issues and search issues are related. The success of users' search
interactions depends on the intersection of more effective search tech-
niques and self–user training.

CONCLUSION AND FURTHER RESEARCH
    Our ongoing study of Web searching is examining a number of large-
scale Web query transaction logs. These studies, using large-scale log data,
are showing some interesting trends and patterns in general Web search-
ing and helping to answer some interesting questions about Web search-
ing. Due to the nature of the data, the research cannot address the results
of users' queries or assess the performance of different search engines.
However, the findings do provide a snapshot for comparison of public Web

searching that can help improve Web search engines and services. Further research is currently being conducted, using query data from Alta Vista, to explore Web search including the similarities and/or differences between North American and European users. Ongoing Web user behavior research is further identifying trends and impacting the development of new types of user training, interfaces and software agents, and new organizational schemas to aid users in better 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., 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. (2002). Trends in multimedia Web searching: 1997–2001. *Information Processing and Management, 38*(3), 475–496.

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., & 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., Wolfram, D., & Saracevic, T. (2002). From e-sex to e-commerce: Web search changes. *IEEE Computer, 35*(3), 133–135.

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. (in press). Web searching for sexual information: An exploratory study. *Information Processing and Management.*

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., 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. (in press). Medical and health Web searching: An exploratory study.

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.