
Searching the Web: a survey of EXCITE users

*Amanda Spink
Judy Bateman and
Bernard J. Jansen*

The authors

Amanda Spink is Assistant Professor in the School of Library and Information Science, University of North Texas, Denton, USA.

Judy Bateman is Assistant Professor in the School of Library and Information Science, Catholic University of America, Texas, USA.

Bernard J. Jansen is Executive Officer in the Department of Electrical Engineering, United States Military Academy, West Point, USA.

Keywords

Search engine, Surveys, World Wide Web

Abstract

Web search services are now a major source of information for a growing number of people. We need to know more about how users search Web search engines to improve the effectiveness of their information retrieval. This paper reports results from a major study exploring users' information searching behavior on the EXCITE Web search engine. The study is the first to investigate Web users' successive searching behavior as they conduct related searches over time on the same or evolving topic. A total of 316 EXCITE users responded to an interactive survey accessed through EXCITE's homepage. Users provided information on their search topics, intended query terms, search frequency for information on their topic, and demographic data. Results show that when searching the Web: users tend to employ simple search strategies; and often conduct more than one search (successive searches) over time to find information related to a particular topic. Implications for the design of Web search services are discussed.

Internet Research: Electronic Networking Applications and Policy
Volume 9 · Number 2 · 1999 · pp. 117-128
© MCB University Press · ISSN 1066-2243

Introduction

The Web is a collection of information resources of all kinds. However, the selection, organization, and retrieval of Web materials are not standardized and wide variation exists in the access capabilities of Web search engines. Most Web search engines are similar to access mechanisms to digital libraries and information retrieval (IR) systems that use IR techniques (e.g. Boolean queries and relevance ranking). In the broadest sense, the Web search engines are a gateway to information services including digital libraries and IR. In many cases, users must search the Web to access digital libraries and IR systems – particularly if they do not know specific URLs for these services. Owing to the growing importance of Web searching and the need for more effective Web search services, a growing body of research is beginning to investigate users' interactions with Web search services (e.g. EXCITE). Behavior common to IR systems users can also be investigated with Web users. The aim of the study presented in this paper is to explore if behaviors exhibited by IR system users, i.e. successive or related searches concerning the same or evolving information problem, are also common to users of Web search engines.

Many IR researchers have modeled users' interactions with IR systems. Ingwersen (1992, 1996) took a broad approach and suggested cognitive representations by all participants in interaction, users, texts, intermediaries and systems, to serve as the base for a cognitive model of IR. Belkin *et al.* (1995) took a more specific approach and treated IR interactions as a series of episodes or frames, each of which supports different types of user interactions and tasks. Saracevic (1996, 1997) proposed a stratified model of IR interactions. Saracevic's stratified model of IR is based on Storr's (1994) approach to interaction as a discourse between a user and "computer" through the interface.

The authors gratefully acknowledge the assistance of Graham Spencer, Doug Cutting, Amy Smith and Catherine Yip of EXCITE, Inc. We also thank Mark Wilcox, Leslie Burkett and Nancy Spaid of UNT, and Tefko Saracevic of Rutgers University, and the anonymous reviewers for their contribution to the development of this research.

Web searching may also be characterized as an interactive process that proceeds in a dynamic way at different levels.

An important part of users' interactions with IR systems is their patterns of successive searches over time related to the same or evolving information problem. The next section of the paper discusses what we know about users' successive searches on IR systems.

Successive searching

Humans seek information in stages over extended periods as their information problem changes (Kuhlthau, 1993) and use different types of IR systems during an information seeking process (i.e. Web, CD-ROMs). Recent research in the IR context shows that users with a broader problem-at-hand often seek information in stages over extended periods and use a variety of information resources (Spink, 1996). As time progresses, many users search the same or possibly different interactive systems (digital libraries, IR systems, Web services) for answers to the same or evolving problem-at-hand (Bateman, 1998). The process of conducting related or successive searches over time in relation to a given, possibly evolving information problem (including changes or shifts in beliefs, and cognitive, affective, and situational states) is called successive searching (Spink *et al.*, 1998c).

For example, an MBA student conducts a broad search of the CD-ROM database Management Contents for information on employee representation on company boards. The results of the search include a number of articles that discuss employee representation on the boards of Scandinavian companies. The next day the student refines their topic and searches the CD-ROM version of the ABI/Inform business database for articles on employee representation on the boards of Scandinavian companies. In this case, the user's successive search pattern included: an initial search on the Management Contents database for information on employee representation on company boards; followed by a change from the Management Contents database to the ABI/Inform database, and a more specific search and modification to the original search strategy with the addition of the new concept – Scandinavian. The user may also enter the name of specific Scandinavian countries or companies derived from the results of

the first search. Recent studies suggest that patterns of related and successive searches may be a fundamental aspect of many users' searching behavior and interaction with IR systems when seeking information related to an information problem (Spink, 1996).

Studies show that many IR system users (Saracevic *et al.*, 1991), end users (Huang, 1992), and OPAC users (Robertson and Hancock-Beaulieu, 1992) conduct successive searches when seeking information related to a particular information problem. Robertson and Hancock-Beaulieu (1992) found a continuity of search topics and relevance judgments by the same OPAC users over successive searches. Some users explored a topic over an extended period and interacted at intervals with the online catalog OKAPI, using identical or closely related search strategies. Studies by Spink (1996), Bateman (1998a) and Spink *et al.* (1998a) found that generally one in two users of IR systems have conducted more than one (1) IR search at different stages of their information seeking process on a particular topic.

Users' successive searching behaviors currently receive little, if any, support from present IR interfaces, procedures, or search engines. By and large, interactive IR systems are built following a single search paradigm, i.e. they are designed and operate on the assumption that every search is unrelated to a user's previous or future searches. A growing body of studies is exploring users' Web searching behavior. However, we know little about users' successive searching behavior on the Web.

Web searching

The phenomenal growth in the size of the Web has created a growing body of empirical research investigating many aspects of users' interactions with the Web. User-oriented Web research generally includes experimental and comparative studies, user surveys, and user traffic studies (Crovella and Azer, 1996). Experimental and comparative studies show little overlap in the results retrieved by different search engines based on the same queries (Ding and Marchionini, 1996; Lawrence and Giles, 1998). Many differences in search engine features and performance (Chu and Rosenthal, 1996) and users' Web searching behavior (Tomaiuolo and Packer, 1996) have been

identified. A growing number of studies are comparing novice and expert Web searchers (Hoelscher, 1998). Some studies have found regular patterns in Web users' surfing behavior (Huberman *et al.*, 1998).

Many surveys of Web users have been conducted, either library based (Tillotson *et al.*, 1995) or distributed via newsgroups (Perry, 1995). Pitknow and Kehoe (1996) found major shifts in the characteristics of Web users over four surveys, including a growing diversity of Web users based on age, gender, and access through both the office and home computers. Hoffman and Novak (1998) found that white and African Americans use the Web differently. Whites were twice as likely to use the Web and have a computer in their home than African Americans.

However, few studies have examined the information problems that underlie users' interactions with the Web or the number of Web searches users conduct to find information on a particular topic. This paper reports results from the first study to investigate the number of Web searches users conduct to find information on a particular topic when using the EXCITE search engine. The research reported in this paper is significant because, as the size and variety of information resources on the Web grow exponentially, the problem of searching becomes a critical problem for a growing number of Web users. The design of Web search engines is generally driven by technological criteria and technology-related algorithms, and is found lacking in many respects when evaluated by users. The aim of the study was to derive human dimensions and criteria for the design of improved Web interfaces and search engines.

The next section of the paper outlines the research objectives of the study.

Research objectives

The major objective the study reported in this paper was to gather data on the use of a major Web search engine to begin to model Web user characteristics and search behaviors. Specifically, data were collected examining users':

- demographic characteristics;
- search topics;
- search terms and queries;
- successive search behavior.

Research design

Data collection and analysis

Users of the EXCITE Web search service were asked to complete an interactive survey form about their interaction with EXCITE, including questions to determine users' current search topic, search terms, information seeking stage, and frequency of searching EXCITE on their current topic. Limitations of this study include the small sample size, the exclusive use of an interactive survey form and the dependence on respondents' self-reported behaviors. An interactive survey form of 18 questions was developed by the researchers in collaboration with EXCITE, Inc. (see Appendix) using previous Web surveys by Pitknow and Kehoe (1996) and a survey of IR system users' successive searching behavior (Spink, 1996). The survey instrument was pilot tested with students from the University of North Texas. The survey was made available through EXCITE's home page (<http://www.excite.com>) for five days from Friday, April 11 to Tuesday, April 15, 1997. Users accessing EXCITE through a meta-search tool or another service could not access the survey form. Respondents clicked the "Send Survey" button at the completion of the survey. The survey drew 11,187 http requests (approximately 3,729 visitors) over the five-day period. A total of 480 respondents or approximately 7.7 percent of site visitors clicked the "Send Survey" button at the end of the survey form. The period of heaviest usage of the survey form was 10am to 2pm on Saturday, April 12. In 12 questions, respondents selected one answer from a number of options; in two questions, respondents chose either "Yes" or "No"; in one question either "Male" or "Female"; and in three questions, respondents also provided written descriptions of their current search topic, listed their proposed search terms and provided comments on their searching or on the survey. The written responses were qualitatively analyzed and the responses categorized. The survey data were tabulated using the ACCESS statistical package. Technical difficulties also resulted in some data corruption on the server for five questions.

Results

This paper extends previous analysis reported at the WebNet 98 conference (Spink *et al.*, 1998b). First a brief overview of the demographic data is provided. More detailed analysis is provided for respondents' search topics, search terms and queries, and successive searching. A total of 316 (75.2 percent) of the "Sent" survey forms contained usable data. One respondent returned 50 blank survey forms in a row. Some respondents did not provide answers to all survey questions, but their responses were included where applicable. The next section of the paper briefly outlines respondents' demographic characteristics.

Demographic characteristics

Age

Respondents ranged in age from less than ten years to over 60 years, with two peaks in the data; the first was respondents aged 21-30 and the second group of respondents were aged 41-50; the majority of respondents were between the age of 20 and 50 years (Table I) – not quite the X-generation.

Education level

Most respondents were either high school or college graduates (Table II).

Occupation

Students and professionals formed the largest group of respondents, followed by executives and the self-employed (Table III). Overall, many respondents were from business or academic related environments. It is not surprising that the college crowd formed a large group of respondents. The occupational categories are not mutually exclusive, as many respondents could probably classify themselves under more

Table I Age of survey respondents

Age (years)	Number	Percentage
Under 10	4	1
11-20	48	16
21-30	69	23
31-40	58	19
41-50	62	21
51-60	36	12
61+	24	8
Total	301	100

Table II Education level of respondents

Education level	Number	Percentage
High school	57	19
Vocational	24	8
Some college	71	24
Bachelor's	70	24
Master's	46	16
Professional	14	5
Vocational	24	8
PhD	9	3
Student	2	1
Total	293	100

than one category, e.g. a marketing manager who is also an MBA student.

Unfortunately, the respondents' gender data were lost and we were unable to compare gender, education, age and occupation.

Computer domain

Interestingly, the largest group of respondents were searching EXCITE from home – followed by commercial and educational users (Table IV). However, we do not know how many respondents were searching both at home and at work.

Geographic location

The overwhelming number of respondents were located in the USA with a smaller number of respondents from Western Europe, the UK and the rest of the world (Table V). This finding was not unexpected and reflects the current concentration of Web searching in the USA. The

Table III Occupation of respondents

Occupation	Number	Percentage
Student	55	19
Professional	38	13
Executive	29	10
Self-employed	25	9
Technical	16	6
Faculty/academic	14	5
Consulting	14	5
Services	14	5
Research and development	11	4
Clerical	10	4
Marketing	8	3
Other	50	18
Total	284	100

Table IV Computing domain of respondents

Computing domain	Number	Percentage
Personal	107	36
Commercial	83	28
Educational	55	18
Organizational	22	7
Government	7	2
Military	2	1
Other	23	8
Total	299	100

survey was also only available in English, which may have further restricted the number of respondents.

Overall, respondents were North Americans aged 20-50 years old with a college education who were searching EXCITE via an IBM PC from their office, home or school.

The next section of the paper shows the topics respondents were searching for on EXCITE.

Search topics

Respondents were asked to provide a written description of their current search topic. A qualitative analysis of respondents' search topic descriptions yielded 16 broad search topic categories. Table VI lists the frequency of searches within the 16 search topic categories.

Table V Geographic location of respondents

Geographic location	Number	Percentage
North America	225	84
Western Europe	8	3
UK	5	2
Middle East	3	1
Latin America	6	2
Africa	3	1
South East Asia	4	1
South Asia/India	3	1
Central America	3	1
Australia	2	1
Japan	2	1
Eastern Europe	2	1
Korea	1	1
Other	1	1
China	1	1
Total	269	100

Table VI Frequency of search topics

Search topic category	Number of searches	Percentage of searches
<i>Individual or family</i>		
Family or friend	17	6
Public figure	11	4
Genealogy	6	2
Sub-total	34	12
Computers	34	12
Business	30	10
Entertainment	23	8
Medical	22	8
Politics and government	20	7
News	19	7
Hobbies	18	6
General information or surfing the Web	16	6
Science	16	6
Travel	13	5
Arts and humanities	12	4
Education	10	3
Shopping	8	3
Graphic images	7	2
Employment	5	1
Total	287	100

Respondents' searches were dispersed over a broad range of topics that were similar in many ways to public library general reference questions. The major topic of EXCITE searches was for information about people, companies or computer products. Interestingly, respondents did not report any topics related to adult interest. As Jansen *et al.* (1998) found that sex was a major topic of EXCITE users' search queries, we concluded that EXCITE users seeking sexually oriented information were not willing to complete our survey form. Despite the lack of sexually oriented queries we might have expected we were able to determine some important characteristics of users' searching behavior:

- Most respondents searched on a single topic as determined by their query terms and search topic statements.
- *Multiple search topics* were determined by an analysis of the query terms and search topic statements. Eleven respondents reported searching on two different topics and two respondents searching on three topics. The topics for respondents who reported

browsing or surfing, or as one respondent put it “whatever interests me”, were categorized as general information or surfing searches.

The next section of the paper reports on users’ search terms and queries.

Search terms and queries

Table VII provides a more detailed overview of the search terms reported by respondents.

These are not necessarily the terms or phrases that respondents used when searching, but the terms that respondents reported were those they intended to use, not those actually used. We were not able to obtain the data from EXCITE to compare intended and actual queries. We did find that:

- The mean number of search terms was relatively low at 3.34 per search.
- Some respondents seemed confused about what they were to report when asked to list query terms for their search.

Some respondents reported links instead of query terms and six respondents used the query term area to describe their search. One respondent put question marks in the query term area. Use of Boolean operators and search modifiers was also an area for examination.

Boolean operators and search modifiers

EXCITE allows searching for phrases, Boolean operators (AND, OR, and AND NOT), and uses parentheses to group search terms and

Boolean operators. Many respondents included terms that they clearly meant as a phrase or proper name, but no respondent indicated that they would use quotes (EXCITE’s method of indicating that two or more words should be next to each other) around these phrases. EXCITE also allows the user to mark words with a “+” (plus) to indicate that the retrieved information must contain this word. A “-” (minus) is used to indicate that the retrieved information must not contain that word. Terms are searched as a phrase only when the phrase is enclosed in quotation marks, i.e. “endangered species”. If a phrase is entered without the quotation marks terms will be connected by the Boolean OR operator, i.e. endangered species without quotation marks will result in a query of endangered OR species. Our analysis showed that:

- Some respondents reported the format and syntax of their search queries in addition to the search terms that they planned to use.
- Few queries included Boolean or other operators. Of the ones that did:
 - four queries included AND;
 - two queries included OR; and
 - 11 queries included +.

One respondent used both AND and OR and parentheses in their search query. This respondent also attempted to truncate using an asterisk (*). EXCITE does not use an asterisk as a truncation operator so the query would retrieve information that contained the word stem followed by an asterisk, i.e. *librar** would retrieve only *librar** and not library or libraries. EXCITE help facilities do not mention a truncation operator.

- An additional seven respondents used the word “and” in a manner that indicated they were intending it as the Boolean AND operator. EXCITE requires that AND be capitalized to be considered a Boolean operator; otherwise it will be treated as a stop word.
- Respondents used both “and” and AND to connect words that they seemed to think would be automatically searched as phrases. Without the quotation marks each term in the phrase is automatically combined with an implicit Boolean OR.
- Some respondents used the “+” sign instead of the Boolean AND. Since a “+” is used to

Table VII Search term data

Classification	Number of search terms
Total number of respondents who reported terms	210
Total terms (did not include stop words)	701
Mean number of terms/respondent	3.34
2-term phrases	84
3-term phrases	9
Proper nouns (names, places, companies, etc.)	45
Links	9
Described search	6
URL	1

indicate that the retrieved information must contain this word it can be used in place of the Boolean AND operator. However, the initial term must also be preceded with a “+” for the query to have the same results as an AND operator.

- Five respondents used the “+” correctly and placed it in front of the desired word with no space between the “+” and the word. Two respondents incorrectly added a space.
- A total of 24 (9 percent) respondents used Boolean operators, “+” signs or “and” in a manner that indicated that they expected it to be a Boolean operator.
- A total of ten respondents used the correct syntax for EXCITE in their search queries.
- No respondent used a “-”, quotation marks, or the Boolean operator AND NOT.

Overall, few respondents were planning to employ Boolean operators and even fewer users applied the correct syntax to enter search phrases and Boolean operators.

Results from a separate analysis of 51,000 EXCITE user queries by Jansen *et al.* (1998) support this low use of Boolean operators; only one in nine queries contained Boolean operators or search modifiers. EXCITE uses the Boolean OR as a default operator that can result in searches that are less specific than the user intended and an increase in the search’s retrieval. EXCITE ranks and posts retrieved information by relevance ranking and this may help compensate for incorrect search query syntax. However, when systems calculate relevance rankings usually both proximity and frequency of terms are considered. The user who thinks he or she is searching a phrase by simply entering the terms into the search statement in phrase order may obtain results that have high relevance rankings but do not relate well to the user’s intended search query. Web search services may consider providing more help features or interactive tutorials to educate and guide the users through the correct application of Boolean operators and modifiers.

The next section of the paper reports results from the questions related to respondents’ successive searching behavior when using the EXCITE search engine.

Successive searching behavior

Frequency of EXCITE searching

Respondents were first asked how frequently they searched EXCITE in general for information (Table VIII).

Many respondents reported searching EXCITE either for the first time or on a daily basis to find information, and nearly a third of respondents were also searching EXCITE weekly or at least two to three times per week. It appears that our respondents were in large part frequent EXCITE users.

Respondents were then asked to estimate the number of EXCITE searches they had conducted on their current topic.

Number of EXCITE searches on current topic

As Table IX shows, 40 percent of respondents were conducting their first search of EXCITE on their current topic; 31 percent of respondents reported a pattern of successive searches of between one to five EXCITE searches on their current topic; 30 percent reported more than five EXCITE searches on their topic; and 13 percent reported conducting

Table VIII Frequency of EXCITE searching

Frequency of searches	Number of users	Percentage of users
First search	51	17
Daily	124	42
2-3 searches	57	20
Weekly	45	15
Monthly	19	7
Total	292	100

Table IX Number of EXCITE searches on current topic

EXCITE searches	Number of users	Percentage of users
First search	112	39
2-5	88	31
6-10	32	11
11-15	10	3
16-20	9	3
20+	37	13
Total	288	100

more than 20 searches on their topic. By respondent estimates, many users had repeatedly searched EXCITE over time for information on the same or evolving topic.

Relevant retrieval

Respondents were then asked if they had retrieved any relevant information from EXCITE on their current topic. Most respondents reported retrieving relevant information from EXCITE on their current topic (Table X).

Information seeking stages

Respondents were then asked to assess their current information seeking stage related to their current search topic as:

- (1) beginning information gathering;
- (2) still information gathering; or
- (3) completing information gathering (Table XI).

Different EXCITE respondents were at different stages of their information seeking process related to their current search topic. Most respondents reported that they were:

- (1) still gathering information on their topic (50 percent), and
- (2) conducting successive searches of EXCITE or frequently searching for information over time during an information seeking process related to a specific search topic (61 percent).

Table X Users' retrieval of relevant information from EXCITE on current topic

Retrieval status	Number	Percentage
Yes	206	72
No	80	28
Total	286	100

Table XI Stage of respondents' information gathering on their current topic

Stage of information gathering	Number of users	Percentage of users
Beginning	112	39
Still gathering	141	50
Completing	31	11
Total	284	100

Number of searches by information seeking stage

Constructed from a combination of Table IX and Table XI, the matrix Table XII shows that respondents were at different stages of their information seeking process and had conducted successive searches when seeking information on a particular search topic.

The largest group of EXCITE respondents (23 percent) were conducting their first search at the beginning of their information seeking process on their current topic. Some 26 (10 percent) respondents also reported still gathering information after more than 20 EXCITE searches. The largest group of respondents had conducted from one to five searches, many at the beginning or still gathering stages of their information seeking process.

Changes in search terms

A total of 54 percent of respondents, who had conducted more than one EXCITE search on their current topic, reported changing their search terms over successive searches (Table XIII).

However, the other half of successive searchers reported "still gathering" or "completing" with no change in their search terms over successive searches. This finding was not surprising, as previous studies reported similar findings with IR system, CD-ROM and Online Public Access Catalog (OPAC) users (Robertson and Hancock-Beaulieu, 1993; Spink, 1996; Spink *et al.*, 1998a). Successive searching involves potential changes and shifts in many

Table XII Matrix of users' information gathering stage by number of times EXCITE searches on current topic

EXCITE searches	Beginning stage	Still gathering	Competing information gathering
First	63 (23)	35 (13)	7 (3)
2-5	23 (8)	48 (18)	11 (4)
6-10	8 (3)	16 (6)	7 (2)
11-15	2 (1)	6 (2)	2 (1)
16-20	2 (1)	4 (1)	2 (1)
20+	6 (2)	26 (10)	4 (1)
Total (272 users)	104 (38)	135 (50)	33 (12)

Note: Numbers in parentheses are percentages

Table XIII Change in users' search terms on current topic

Status	Number	Percentage
Yes	138	54
No	119	46
Total	257	100

variables, such as changes in search terms, search strategies, relevance judgments and criteria, or information problem focus. Those respondents who had conducted successive searches were asked if their search terms had changed over successive searches. The study did provide a rich set of data and some surprising findings that are discussed further in the next section of the paper.

Discussion

The results of the study revealed a number of interesting findings:

- EXCITE users seem to be a mixed bunch. They span not only most age groups, but also different educational and occupational backgrounds ranging from academia to business.
- They seem to prefer to access the Web via IBM PCs.
- They are largely scattered across North America.
- Respondents' search topics varied immensely, from entertainment to business and computing. The topics were similar to reference queries that might be made to a reference librarian on a public library. Recent research found sex to be the most frequent search topic during an analysis of over 51,474 EXCITE search queries from over 18,113 EXCITE users (Jansen *et al.*, 1998; Jansen *et al.*, 1998a,b). In light of this recent study of EXCITE queries, the lack of sexually motivated search topics and terms reported by survey respondents was rather surprising and probably due to self-censorship by EXCITE users.
- Respondents were not proposing to use many search terms or employ complex search strategies. Nor were they planning to use many search features, such as Boolean operators, query modifiers or natural language queries. This finding implies a fairly low level of interaction with the EXCITE Web search

engine. This finding does not account for respondents' actual behavior on EXCITE, but it does give some insight into their initial search terms and strategies.

- Many respondents were conducting successive searches on their topic. Users were repeatedly conducting related searches over time on the same topic. The sheer magnitude of any retrieval in response to a few search terms may cause users to quickly peruse the results, logoff, possibly rethink or search another information resource, and then have another go on the EXCITE search engine.

A recent study found that EXCITE users performed limited query reformulation and had little persistence in viewing retrieved lists of Web sites (Jansen *et al.*, 1998; Jansen *et al.*, 1998a,b). Overall, the users' ability to specify good search terms and create complex search queries to clearly and precisely capture relevant retrieval seems rather low. EXCITE users also appear to lack the motivation to employ complex search strategies and learn correct syntax and rules, and may expect the search engine to automatically create effective queries.

The results of this study show that successive search episodes are an important behavior for many users of Web search engines. Successive searches are related searches during the same or evolving information problem. We can speculate as to why users conduct successive searches over time related to the same or evolving information problem. A user's first search may not bring complete or satisfactory results. After time spent examining results from the first search a user may decide to revise their search strategy and terms during a second search or subsequent searches on different databases or Web search engines. Particularly for users with more complex information problems, e.g. information on how to build an addition to their house, as their understanding or learning related to that information problem develops they may conduct many searches on different databases or Web search engines over time. Some users with an ongoing and specific information need, such as tracking the appearance of new Web sites related to model airplanes, may enter the same search terms over time to keep up-to-date. Further studies are needed to examine why Web users conduct successive searches and the

patterns of successive searching over time, including the changes or shifts in successive search episodes. The results of this study have implications for the design of Web search services.

Implications for Web design

It seems fairly obvious that allowing users to save and reuse their search terms, strategies and results for later reformulation should be a feature of Web search interfaces and systems. After all, online database services such as DIA-LOG have provided users with "save search" features for many years. In fact many other features commonly provided by online database systems are not currently incorporated into Web search engines, e.g. keyword in context, save searches, many types of proximity operators, database selection tools, interactive information sheets about databases, etc. Using a meta search tool is more effective than using single Web engines search (Selberg and Etzioni, 1995; Gauch *et al.*, 1996; Chang and Hsu, 1997; Dong, 1998; Lawrence and Giles, 1998) and may allow more effective saving and reformulating during successive searches. As we move from single search engines to meta search tools, such as Dog Pile, Metacrawler and Inquirus (Lawrence and Giles, 1998), the ability to save searches and results may be more feasible to implement. Users cannot currently transfer searches from one Web search engine to another.

Conclusions and research

Users have the option to engage in fairly complex processes with Web search engines and use their full functionality to improve their retrieval results. However, most Web searches are short and simple. This paper has identified a crucial problem for search engine designers – the lack of transparency of both the nature and benefits of basic and advanced search features for the large mass of users who frequently interact with the Web. Many users are successively searching the Web. Web search engines and interfaces do not effectively support this behavior. Further research is needed to explore the relationship between users' topic complexity, their information seeking behaviors and

related Web successive searching. Research is currently modeling the changes and shifts that occur within and between users' successive searches on IR systems (Spink *et al.*, 1998c). Research is needed to model successive searching behavior by Web users, including the shifts and changes during and between their successive searches.

Research by the same group of researchers is also analysing 51,000 user queries provided by EXCITE Inc. (Jansen *et al.*, 1998; Jansen *et al.*, 1998a,b) and a larger data set of 1 million EXCITE users' queries.

References

- Bateman, J. (1998), *Changes in Relevance Criteria: An Information Seeking Study*, unpublished PhD Dissertation, University of North Texas, School of Library and Information Sciences.
- Belkin, N.J., Cool, C., Stein, A. and Theil, U. (1995), "Cases, scripts and information seeking strategies: on the design of interactive information retrieval systems", *Expert Systems with Applications*, Vol. 9 No. 3, pp. 379-95.
- Chang, C-H. and Hsu, C-C. (1997), "Customizable multi-engine search tool with clustering", *Proceedings of the Sixth International World Wide Web (WWW6) Conference*, April 7-11, Santa Clara, CA.
- Chu, H. and Rosenthal, M. (1996), "Search engines for the World Wide Web: a comparative study and evaluation methodology", *Proceedings of the 59th Annual Meeting of the American Society for Information Science*, Baltimore, MD, pp. 127-35.
- Crovella, M.E. and Azer, B. (1996), "Self-similarity in World Wide Web traffic evidence and possible causes", *Proceedings of ACM SIGMETRICS*, Philadelphia, PA, May, pp. 126-37.
- Ding, W. and Marchionini, G. (1996), "A comparative study of Web search service performance", *Proceedings of the 59th Annual Meeting of the American Society for Information Science*, Baltimore, MD, pp. 136-42.
- Dong, J. (1998), *Combination of Multiple Web Search Results and its Effects on the Search Performance*, unpublished paper, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign.
- Gauch, S., Wang, G. and Gomez, M. (1996), "Profusion: Intelligent Fusion from Multiple, Distributed Search Engines", <http://www.tisl.ukans.edu/~sgauch/papers/JUCS.html>

- Hoelscher, C. (1998), "How Internet experts search for information on the Web", *Proceedings of WebNet 98*, Orlando, FL, November 8-12.
- Hoffman, D.L. and Novak, T.P. (1998), "Bridging the racial divide on the Internet", *Science*, April 17, pp. 390-1.
- Huang, M.H. (1992), *Pausing Behavior of End-users in Online Searching*, unpublished dissertation, University of Maryland.
- Huberman, B.A., Pirolli, P.L.T., Pitkow, J.E. and Lukose, R.M. (1998), "Strong regularities in World Wide Web surfing", *Science*, April 3, pp. 95-7.
- Ingwersen, P. (1992), *Information Retrieval Interaction*, Taylor Graham, London.
- Ingwersen, P. (1996), "Cognitive perspectives of information retrieval interaction: elements of a cognitive IR theory", *Journal of Documentation*, Vol. 52 No.1, pp. 3-50.
- Jansen, B.J., Spink, A. and Saracevic, T. (1998a), "Failure analysis in query construction: data and analysis from a large sample of Web queries", *Proceedings of the 3rd ACM Digital Libraries Conference*, June 23-26, Pittsburgh, PA, pp. 289-90.
- Jansen, B.J., Spink, A. and Saracevic, T. (1998b), "Searchers, the subjects they search and sufficiency: a study of a large sample of EXCITE searches", *Proceedings of the WebNet 98 Conference*, November 9-13, Orlando, FL.
- Jansen, B.J., Spink, A., Bateman, J. and Saracevic, T. (1998), "Real life information retrieval: a study of user queries on the Web", *SIGIR Forum: Association of Computing Machinery - Special Interest Group on Information Retrieval Research and Development*, ACM, New York, NY, Vol. 32 No. 1, pp. 5-17.
- Kuhlthau, C. (1993), *Seeking Meaning: A Process Approach to Library and Information Services*, Ablex Press, Norwood, NJ.
- Lawrence, S. and Giles, C.L. (1998), "Searching the World Wide Web", *Science*, April 3, pp. 98-100.
- Perry, C. (1995), "Travelers on the Internet: a survey of Internet users", *Online*, March/April, pp. 29-34.
- Pitkow, J.E. and Kehoe, C.M. (1996), "Emerging trends in the WWW user population", *Communications of the ACM*, Vol. 39 No. 6, pp.106-8.
- Robertson, S.J. and Hancock-Beaulieu, M. (1992), "On the evaluation of IR systems", *Information Processing and Management*, Vol. 28 No. 4, pp. 457-66.
- Saracevic, T. (1996), "Modeling interaction in information retrieval (IR): a review and proposal", *Proceedings of the 59th Annual Meeting of the American Society for Information Science*, Vol. 34, pp. 3-9.
- Saracevic, T. (1997), "Extension and application of the stratified model of information retrieval interaction", *Proceedings of the 60th Annual Meeting of the American Society for Information Science*, Vol. 35.
- Saracevic, T., Mokros, H., Su, L. and Spink, A. (1991), "Interaction between users and intermediaries in online searching", *Proceedings of the 12th National Online Meeting*, May 12, New York, NY, pp. 329-41.
- Selberg, E. and Etzioni, D. (1995), "Multi-service search and comparison using meta crawler", *Proceedings of the 1995 International World Wide Web (WWW3) Conference*. <http://www.w3.org/pub/conferences/WWW4/papers/169>.
- Spink, A. (1996), "A multiple search session model of end-user behavior: an exploratory study", *Journal of the American Society for Information Science*, Vol. 47 No. 8, pp. 603-9.
- Spink, A., Bateman, J. and Greisdorf, H. (1998a), "Successive searching behavior during interactive information retrieval (IR): development of a new line of research", *Proceedings of the 26th Annual Conference of the Canadian Society for Information Science*, June 3-5, Ottawa, Canada, pp. 401-5.
- Spink, A., Bateman, J. and Jansen, B. J. (1998b), "Users' searching behavior on the EXCITE Web search engine", *Proceedings of WebNet 98*, Orlando, FL, November.
- Spink, A., Wilson, T.D., Ellis, D. and Ford, N. (1998c), "Modeling users' successive searching: a National Science Foundation/British Library study", *D-Lib Magazine*, April, CNI.
- Storrs, G. (1994), "A conceptualization of multiparty interaction", *Interacting with Computers*, Vol. 6 No. 2, pp. 173-89.
- Tillotson, J., Cherry, J. and Clinton, M. (1995), "Internet use through the University of Toronto: demographics, destinations and users' reactions", *Information Technology and Libraries*, September, pp. 190-8.
- Tomaiuolo, N.G., and Packer, J.G. (1996), "An analysis of Internet search engines: assessment of over 2,000 search queries", *Computers in Libraries*, Vol. 16 No. 6, pp. 58-62.

Appendix: EXCITE Survey (HTML version available at <http://www.unt.edu/survey/excite.html>)

Information behavior of EXCITE users

Research provided by the School of Library and Information Sciences at University of North Texas, home of the Mean Green Eagles.

Thank you for participating in our survey. The results of this survey will be used to help improve EXCITE services:

- (1) How often have you used EXCITE to find information?:
 - First time
 - Daily
 - 2-3 times per week
 - Weekly
 - Monthly
- (2) Describe your current search topic (e.g. Looking for information on health care companies with offices in Minneapolis who ...):
- (3) List your query terms for your current topic (e.g. health care, Minneapolis...):
- (4) Status of your information gathering on your current topic?:
 - Beginning information gathering
 - Still gathering information
 - Completing information gathering
- (5) How many times have you used EXCITE to find information on your current topic?:
 - First time
 - 2-5 times
 - 6-10 times
 - 11-15 times
 - 16-20 times
 - 20+ times
- (6) If you have used EXCITE previously for information on your current topic, have your search query terms changed since your first search of EXCITE?
 - Yes
 - No
- (7) Have you found relevant information on your current topic using EXCITE?
 - Yes
 - No
- (8) What other sources have you used or do you plan to use to find the information on your current topic?:
 - Other Web browsers
 - Databases, e.g. DIALOG
 - Libraries
 - Human sources
 - Print sources, e.g. newspapers, magazines
 - Others
- (9) Comments on any of the above questions or related issues:
- (10) Your age?
- (11) Your gender?:
 - Female
 - Male
- (12) Number of hours per week you use a computer?
- (13) Average number of hours per week you use the Internet?
- (14) Type of Internet Domain you use?
- (15) Your occupation?
- (16) Highest level of education?
- (17) Your location?
- (18) Which platform do you use?

Thank you very much for your time and attention to our survey.

Comments or Questions: www@unt.edu