Successive Searching Behaviour During Information Retrieval (IR) Interaction: Development of a Line of New Research

Amanda Spink, Judy Bateman & Howard Greisdorf
School of Library and Information Sciences
University of North Texas
spink@lis.admin.unt.edu

The study of successive searching, or user's pattern of search sessions in digital environments over time related to the same or evolving information problem, is developing as a new line of investigation. A growing body of studies is beginning to investigate and characterize the successive searching process. This area of research draws together information retrieval (IR) research and information seeking research toward a focus on IR within the context of human information behavior. This paper provides results from the first study examining characteristics of mediated successive searches conducted by search intermediaries for 18 clients with evolving information problems. We identify reasons for the successive searches, the frequency of successive searches, and some characteristics of successive searches.

Introduction

Recent research shows that users with a broader problem-at-hand often seek information in stages over extended periods and use a variety of information resources. As the time progresses, IR users tend to search the same or different systems for answers to the same or evolving problem-at-hand. As they learn or progress in their work, or as they clarify a problem and/or question, or as their situational context changes, users come back to IR systems of various kinds for more searches. The process of repeated, successive search sessions 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 *a successive search phenomenon. Successive search episodes* then become units for observation and analysis. A *search episode* is a user interaction with either a single or multiple digital information systems, i.e., CD-ROM databases, online databases, digital libraries, Web search engine, or OPAC which are separated by a

time period (hours, days, weeks, months) of evaluation of the previous search episode before embarking upon a new search episode. The modeling of users in successive searches is then *successive user modeling.* A key dimension is time, the key variable changes or shifts in successive search episodes over time, and the key constant is the same or evolving information problem. The evolution, if any, of a problem and other cognitive, affective and situational variables can be mapped, and the history of successive search episodes can be researched.

However, users' successive searching currently receive little, if any, support from present IR interfaces and procedures, or from Web search engines. By and large, IR systems are built following a single search paradigm, i.e., they are designed and operate on the assumption that every search is an end in itself. Neither the commercial systems (such as Dialog or Lexis/Nexis), nor the TREC (Text Retrieval Conference) (Sparck Jones 1995) experimental IR systems, nor the search engines on the Web, nor the 'intelligent' agents in knowledge base applications, address or support, successive searching. Research in this area is in its formative stage. The study reported in this paper is part of a major research project that seeks to investigate this new and little explored line of inquiry in IR addressing the successive search phenomenon and associated episodes.

Related Research

Information Retrieval Research

Research into the human or cognitive (user modeling) aspects of IR is also in its infancy with a growing body of research on users' interactivity and measures for observing user interactivity (Saracevic, Mokros, Su and Spink 1991). Major theorized interactive IR models have recently emerged but have yet to be developed sufficiently to be empirically tested (Belkin, Cool, Thiel and Stein 1995; Ingwersen 1996; Saracevic 1997). IR researchers are also beginning to investigate the context of users' searches and evaluation (Ellis, 1997) and identify key elements in users' searching within the context of users' single search of an IR system. However, this research has overlooked any successive searches related to the same information problem. Important elements within single searches have been identified, including feedback types, search term selection and elicitations

(Saracevic, Kantor, Trivison and Chamis 1988; Saracevič, Mokros, Su and Spink 1991; Spink and Saracevic 1997). The differences between end-user and intermediary searching behaviors have also been investigated (Hsieh-Yee 1993). Research has also shown that end-users perform different search sessions over time (related to a different information problem) including searches of successive databases or IR systems (Saracevic, Mokros, Su and Spink 1991). In addition, recent findings from general information behavior studies also have strong implications for IR research and the need to investigate users' successive searching behavior.

Information Seeking Studies

Research shows that 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 (e.g., Web, CD-ROM databases). However, recent studies also show that users also conduct successive searches over time when seeking information to solve an information problem (Robertson and Hancock-Beaulieu 1992; Spink 1996). However, limited knowledge exists on the patterns of users' successive search behavior. Results from information seeking studies support the notion of the successive search phenomenon by showing that humans progress through a series of stages, adopt different strategies and exhibit different information behaviors at different stages of their information-seeking process (Ellis 1989; Kuhlthau 1993). Kuhlthau (1993) found that the information search process of library patrons occurred in six clearly defined stages related to the cognitive, affective states and search activities of the users. Although Kuhlthau did not investigate the use of IR systems by library patrons, her findings suggests that IR system users continue to collect and seek information throughout their information-seeking process using or requiring different types of information, conducting different types of searches, and using different search terms and strategies at different stages of an information seeking process (Kuhlthau, Spink and Cool 1992). However, the nature of successive searches and related relevance judgments during a user's longitudinal information search process is in the early stages of development and presents important implications for the design of effective IR systems to support effective user searching. Recent research (Spink and Greisdorf 1997; Spink, Greisdorf and Bateman, in review) suggests that users' partial rel-

evance judgments also play an important role in the search process and are linked to shifts in the user's information problem during single and successive searches. Findings from this study also suggest that partial or less highly relevant items may be "best" for allowing a successive search user to shift or change their information problem or change stages during the early stages of their information search process. Further research is needed to explore the role of users' "high" and "partial" relevance judgments during an information search process including successive searches.

Successive Searching Studies

Interactive Information Retrieval

Data from several recent studies highlights the weakness of research based on the single search approach and the need for studies that classify and categorize users successive search behavior. Recent studies show users conduct successive IR searches when seeking information related to a particular information problem. In one recent study 18 (45%) of academic users were found to have a previous mediated online search on the same topic, frequently with the same search intermediary (Saracevic, Mokros, Su and Spink 1991; Spink 1993). Huang (1992) studied 44 end-users conducting online searches and found 19 end-users conducted successive searches. Robertson and Hancock-Beaulieu (1992) also identified successive searches by users of the OKAPI online catalog. They found a continuity of search topics and relevance judgments by the same users over successive searches as some users were exploring a topic over an extended period and were interacting at intervals with the online catalog, using identical or closely related search strategies. This research highlights the need for longitudinal research at a problem-level of analysis as opposed to a single search level of analysis of searching behavior.

A recent study of 200 academic CD-ROM and Online Public Access Catalog (OPAC) end-users by Spink (1996) shows that successive IR searches are a fundamental aspect of users' behavior when seeking information related to an information problem. End-users' working on research project or paper were asked to estimate how many search sessions they had conducted on a particular information problem. Table 1 from Spink (1996) shows that 113 (56.5%) end-users reported

conducting more than one IR search and 43 (21.5%) end-users reported five or more IR search sessions.

Table 1. Estimated number of search sessions conducted by end-users during their research project prior to the survey interview.

Number of Search Sessions	End-Users		
	Number	%	Cum%
1 search session	87	43.5%	43.5%
2 search sessions	26	13.0%	56.5%
3 search sessions	31	15.5%	72.0%
4 search sessions	13	6.5%	78.5.%
5 search sessions	18	9.0%	87.5%
6 search sessions	7	3.5%	91.0%
7 search sessions	4	2.0%	93.0%
8 search sessions	1	0.5%	93.5%
10 search sessions	2	1.0%	94.5%
11 search sessions	2	1.0%	95.5%
12 search sessions	1	0.5%	96.0%
20+ search sessions	8	4.0%	100.0%
Total	**200**	**100.0%**	

Table 2 shows that many end-users reported conducting successive search sessions at different stages of their information seeking process related to a particular information problem.

Table 2. Current information seeking stage and number of search sessions reported by end-users prior to the survey interview.

Current Information-Seeking Stage	Number of Previous Search Sessions					
	1 - 3	4 - 6	7 - 10	11+	Total	%
Stage 1	39	18	3	2	62	55%
Stage 2	4	0	1	1	6	5%
Stage 3	8	8	1	3	20	18%
Stage 4	4	0	0	2	6	5%
Stage 5	11	5	1	2	19	17%
Total	**66**	**31**	**6**	**10**	**113**	**100%**

Bateman (1998) investigated the search sessions over time conducted by 35 end-users and found that 33 (94%) end-users conducted more than one search session during their information seeking process related to a particular information problem. Table 3 shows the number of successive search sessions conducted by end-users during their information seeking process.

Table 3. Number of successive search sessions conducted by end-users during their research project.

Number of Successive Searches	Number of Respondents	% of Respondents
0 successive searches	1	2.8%
1 successive search	1	2.8%
2 successive searches	10	28.6%
3 successive searches	9	25.8%
4 successive searches	7	20.0%
5 successive searches	4	11.5%
6 successive searches	1	2.8%
7 successive searches	2	5.7%
Total	35	100%

Web Searching

A recent survey conducted by Spink, Batemen and Jansen (1998) found that EXCITE Web search engine users reported conducting successive search sessions over time related to a particular information problem (Table 4).

Table 4. Matrix of current information seeking stage by number of EXCITE searches (272 users).

Matrix of Current Information Seeking Stage By Number of EXCITE Searches			
EXCITE Searches	No. of Users at Beginning Stage	No. of Users Still Gathering	No. of Users Completing Information Gathering
First search	63 (23%)	35 (13%)	7 (3%)
2-5 searches	23 (8%)	48 (18%)	11 (4%)
6-10 searches	8 (3%)	16 (6%)	7 (2%)
11-15 searches	2 (1%)	6 (2%)	2 (1%)
16-20 searches	2 (1%)	4 (1%)	2 (1%)
20+ searches	6 (2%)	26 (10%)	4 (1%)
Total	**104 (38%)**	**135 (50%)**	**33 (12%)**

Previous studies show that users of IR systems and the Web reported performing successive searches related to the same or evolving information problems over time. These studies have identified the basic phenomenon of successive searching. The aim of the study reported in this paper is to provide insight into the characteristics, frequency and reasons why successive searches are conducted.

Research Objectives

The overall purpose of this emerging line of research is to investigate the nature, manifestations, and behavior of the successive search phenomenon in IR, and to derive criteria of use in design of IR interfaces and search engines supporting successive search episodes. The research questions addressed by the study reported in this paper were:

1. What is the frequency of successive mediated searches during a user's information seeking process?

2. Why do users seek successive mediated searches during their information seeking related to a particular information problem?

3. What are some of the characteristics of successive mediated searches related to a particular information problem, e.g., changes in search terms, databases?

Such research is the logical next step for IR research to improve IR and Web system and interface design, and user education.

Research Design

Data Collection and Analysis

Forty novice search intermediaries were recruited from the students at the School of Library and Information Sciences at the University of North Texas during the academic year 1996-97. Each search intermediary had been trained in the DIALOG Information Service. Each search intermediary then recruited a volunteer client seeking information on a particular problem. The search intermediary then worked with the client over one semester to conduct as many searches as necessary to assist the client in resolving their information problem. Each Search intermediary kept a written diary of the entire process, including the search logs, search request form, clients' relevance judgments, working notes, and a report of their search terms, strategy and search interactions. Eighteen (18) of the forty (40) mediated searches (45%) involved successive searches. The report and data from each search intermediary was analyzed qualitatively and quantitatively.

Results

Frequency of Successive Searches

Table 5 shows the number of successive searches requested by the forty clients.

Table 5. Frequency of successive searches.

Number of Successive Search	Number of Clients	% of Clients
0 successive searches	22	55.0%
1 successive search	10	25.0%
3 successive searches	3	7.5%
4 successive searches	3	7.5%
5 successive searches	2	5.0%
Total	**40**	**100%**

Twenty-two (22) (55%) clients requested one search on their information problem and 18 (45%) users requested more then one search on their topic.

Table 6 lists the search topics explored and the number of searches requested by those 18 clients who requested successive searches. A total of 51 searches were performed for clients who requested successive searches - including 18 first searches and 33 following searches. A mean of 2.8 searches were performed per client with a range of 2-5 searches. Clients most frequently requested 2 searches during their information seeking process.

Table 6 also shows that the search topics ranged across the physical science, social sciences, humanities and medical issues. The reasons reported by the search intermediaries for conducting the successive searches centered on the need to refine or extend the first search, based on the client's evaluation of the results of the first search and changes in the clients information problem, by searching different databases or using different search terms, to find more information.

Table 6. Search topics of 18 clients requesting successive searches.

Search Number	Search Topic	No. of Searches	Reasons for Successive Searches
Client 1	Children with A.D.D	3	Too many hits in the previous search - need to refine the search/ Lost data from the previous search/ Previous search did not include all search terms/ Search terms added after initial search
Client 2	Spider varicose veins	2	First search retrieved titles only/ Second search to retrieve abstracts
Client 3	Digital imaging	2	Use elements (company names) from the first search to enhance retrieval in the second search
Client 4	Aspirin/salicylate allergic reactions	2	Prevent information overload during first search/ Allow for off-line relevance judgments from initial search
Client 5	Community/economic development grants	3	Search different databases
Client 6	Taut Satu tribe	2	To secure more valuable information
Client 7	Pascal programing computer language	3	Provide more information on the topic
Client 8	Theatrical or performing arts libraries	2	Client wanted more information

Client 9	Criminal violence	5	Client feedback provided information for later searches/ To adequately gather the information requested by the client/ Client feedback assisted later searches
Client 10	Wood milling equipment	5	Intermediary determined to prove client wrong and exhaust all possibilities
Client 11	Tropical rain forest as a biome	2	Client needs more relevant information
Client 12	Synthesis of a chemical compound	4	Original query terms proved not correct/Original problem statement expanded by client
Client 13	Hepatitis B	2	Increased problem complexity
Client 14	Groundwater pollution by reformulated gasoline	2	Initial search was exploratory to just see what was out there
Client 15	Abused adults	2	Client wanted to see what else was there though satisfied with the initial search
Client 16	Effectiveness of cognitive retraining programs for brain damaged individuals	4	No useable results on databases first searched
Client 17	Health benefits of green tea	4	First search clarified and refined both actual area of interest and query formulation
Client 18	Computer use by the elderly		Refine the search topic

The next section of the paper further investigates the characteristics of successive searches.

Characteristics of Successive Searches

A number of characteristics of the successive searches were investigated, including whether successive searches included the information seeking stage of the clients, the sources of the search terms, and changes in the search terms and databases searched.

Table 7 shows that two-thirds of the client's who requested successive searches reported having conducted no previous information seeking on their topic and one-third had conducted previous information seeking.

Table 7. Information seeking stages of the successive search clients.

Client's Information Seeking Stage	Number of Clients	% of Clients
Beginning stage	12	66%
Previous information seeking	6	33%
Total	**18**	**100%**

Table 8 shows that the clients were the main source of search terms during the successive search process, although intermediaries did contribute search terms in more than 50% of searches.

Table 8. Sources of search terms.

Sources of Search Terms	Number of Clients	% of Clients
Client terms only	8	45%
Search intermediary term only	0	0%
Both client & search intermediary	10	55%
Total	18	100%

Table 9 show that most successive searches for clients involved changes in search terms from the initial search. This is supported by findings by Spink (1996), and Robertson and Hancock-Beaulieu (1993). Most interesting is the use of the same search terms over successive searches during nearly a third of cases.

Table 9. Changed search terms over successive search sessions.

Changed Search Terms	Number of Clients	% of Clients
Yes	15	83%
No	3	17%
Total	18	100%

Table 10 shows that in 50% of cases the databases searched changed over successive searches. However, interestingly the same databases were searched again (with either the same or different search terms) over successive searches.

Table 10. Changed databases over successive search sessions.

Changed Databases	Number of Clients	% of Clients
Yes	9	50%
No	9	50%
Total	18	100%

In 8 (45%) cases successive searches involved a change in both search terms and databases. From the data we can see that successive searches involved change, refinement or extension from the initial search. How the changes evolved depends on the nature of the information problem and the information seeking stages and changes by the client. As suggested by Spink, Greisdorf and Bateman (in review) an important element in changes and shift between successive searches are the client's relevance judgments.

Relevance Judgments

Table 11 shows data concerning the relevance of the retrieved items during searches for 14 clients who requested 14 initial searches and 24 successive searches - data for the other 4 successive search clients was insufficient to be used in this analysis. The mean number of items retrieved per search was 24.4 and the mean search precision was 49.7%.

Table 11. Successive searches (14 clients).

No. of Searches	No. of Clients	Items Retrieved	Mean No. Items Retrieved	No. of Relevant Items	No. of Partially Relevant Items	No. of Not Relevant Items	Mean Precision
First search	14	411 (45%)	29.4	144 (39%)	30 (32%)	237 (50%)	42.30%
Second search	14 (100%)	346 (38%)	24.7	163 (44%)	62 (68%)	121(25%)	65%
Third search	6 (42%)	76 (8%)	12.6	23 (3.5%)	0	53 (11%)	30%
Fourth search	3 (21%)	67 (7%)	22.3	13 (4.5%)	0	54 (11%)	19.40%
Fifth search	1 (7%)	26 (2%)	26	25 (9%)	0	1 (3%)	96.20%
Total	38	926 (100%)	24.4	368 (100%)	92 (100%)	466 (100%	49.70%

An analysis of the data in Table 11 shows that:

1. Mean number of successive searches for the 14 clients was 2.7 with ten (70%) clients requesting more than two searches and 4 (28%) clients requesting more than three searches.

2. Successive searches may not necessarily lead to a decrease in the mean number of items retrieved per search as the number of successive searches increases.

3. Successive searching may not necessarily lead to greater precision as the number of successive searches increases.

4. Successive searches may provide greater clarity for the client, as evidenced by the absence of partially relevant items after second searches.

Discussion

This study has identified some key characteristics of successive searches. A large number of both end-user and mediated searches are successive searches performed at different stages of the information seekers' information seeking process. We found the information seekers, on average, needed two to three searches to satisfy their information seeking process, although a number of information seekers needed more than three searches. These successive searches involved a refinement or extension to the previous searches as new databases were searched and search terms changed based on the evaluation of the results from previous searches due to a change in the information seekers understanding of their information problem—although some successive searches involved no change in databases or search terms.

We did find some interesting results with regard to relevance judgments over successive searches as the precision of successive searches did not increase. However, the absence of items judged partially relevant after the first two searches is intriguing. Six (70%) of the fourteen clients who requested successive searches no longer judged any items partially relevant after the second search. This implies that some users may cluster towards dichotomous relevance judgments—highly relevant and not relevant—after the information seeker has refined their understanding of the their information problem through a number of successive searches. This finding supports Spink and Greisdorf (1997), and Spink, Greisdorf and Bateman (in review) who found a relationship between partially relevant items and changes in information seekers' information problems, and suggest that partially relevant items may be important for users during the early stages of their information seeking process. Data analysis is currently being conducted on another data set of mediated online searches to further investigate this finding.

Conclusion

A great many users engage repeatedly over time in accessing and searching IR systems or the Web resources in relation to the same or evolving problem-at-hand. Successive searching is widely practiced. Yet, the successive search phenomenon has received little research attention, despite the obvious significance of a better understanding of the phenomenon. Thus, this should be considered as an exploratory study concentrating not only on the collection and analysis of data, but also on a number of methodological issues and problems. Methods for the investigation and depiction of shifts within and between successive search episodes is one of these problems. Obviously, there are a number of kinds of shifts on a number of levels. For instance, on the surface level there are semantic, syntactic, and logic shifts in selected search terms and statements; on the situational level there are shifts in problem definition; there are shifts in focus. It is not even clear how various shifts should be classified. In other words, by necessity, shifts and transitions will be a major focus of successive searching research.

A major international collaborative NSF/British Library study between the University of North Texas and the University of Sheffield is currently investigating these issues (Spink, Wilson, Ellis & Ford, 1998). Currently, IR systems and interfaces do not include assistance to users in successive search episodes. This research seeks to derive criteria for systems and interfaces to accommodate users' successive search episodes.

References

Bateman, J. 1998. Changes in users' relevance criteria: An information seeking study. Unpublished Dissertation Proposal. School of Library and Information Sciences, University of North Texas.

Belkin, N. J., Cool, C., Stein, A., & Theil, S. 1995. Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems. *Expert Systems With Applications 9*(3): 379-395.

Ellis, D. 1989. A behavioural approach to information retrieval system design. *Journal of Documentation 45*: 171-212.

Ellis, D. 1997. The dilemma of measurement in information retrieval. *Journal of the American Society for Information Science 47*(1): 23-36.

Hsieh-Yee, I. 1993. Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the American Society for Information Science 44*(3):161-174.

Huang, M. H. 1992. Pausing behavior of end-users in online searching. Unpublished Doctoral Dissertation. University of Maryland.

Ingwersen, P. 1996. Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory. *Journal of Documentation 52*(1): 3-50.

Kuhlthau, C. C. 1993. *Seeking Meaning: A Process Approach to Library and Information Services.* Norwood, NJ: Ablex Publishing.

Kuhlthau, C. C., A. Spink, & C. Cool. 1992. Exploration into stages in the information search process in online information retrieval: Communication between users and intermediaries. *Proceedings of the 55th Annual Meeting of the American Society for Information Science, 29*: 67-71.

Robertson, S. E., & M. M. Hancock-Beaulieu. 1992. On the evaluation of IR systems. *Information Processing and Management 28*(4): 457-466.

Saracevic, T. (1997). Extension and application of the stratified model of information retrieval interaction. *Proceedings of the Annual Meeting of the American Society for Information Science, 34,*

Saracevic, T., H. Mokros, L. Su, & A. Spink. 1991. Interaction between users and intermediaries during online searching. *Proceedings of the 12th Annual National Online Meeting, 12:* 329-341.

Schamber, L. 1994. Human behavior and relevance. *Annual Review of Information Science and Technology 29*: 1-33.

Sparck Jones, K. 1995. Reflections on TREC. *Information Processing and Management 31*(2): 291-314.

Spink, A. 1993. The effects of user characteristics on search outcome in mediated online searching. *Online & CD-ROM Review* 17(5): 275-278.

Spink, A. 1996. A multiple search session model of end-user behavior: An exploratory study. *Journal of the American Society for Information Science 46*(8): 603-609.

Spink, A., J. Bateman, & B.J. Jansen. 1998. Users searching behavior on the EXCITE Web search engine. *Proceedings of the 1998 National Online Meeting, May 12-14, 1998.*

Spink, A., & H. Greisdorf. 1997. Users' partial relevance judgments during online searching. *Online & CD-ROM Review 21*(5): 271-280.

Spink, A., H. Greisdorf, & J. Bateman. (In review). From highly relevant to not relevant: Examining different regions of relevance. *Information Processing and Management.*

Spink, A., & R.M. Losee. 1996. Feedback in information retrieval. *Annual Review of Information Science and Technology 33.*

Spink, A., & T. Saracevic. 1997. Interaction in information retrieval: Selection and effectiveness of search terms. *Journal of the American Society for Information Science 48*(8): 728-740.

Spink, A., T.D. Wilson, D. Ellis, & N. Ford. 1998. Modeling users' successive searches: A National Science Foundation/British Library Study. *D-Lib Magazine* April 1998.