Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2001 Proceedings

International Conference on Information Systems (ICIS)

December 2001

Understanding Successive Searches Across Multiple Sessions Over the Web

Shin-Jeng Lin
LeMoyne College

Follow this and additional works at: http://aisel.aisnet.org/icis2001

UNDERSTANDING SUCCESSIVE SEARCHES ACROSS MULTIPLE SESSIONS OVER THE WEB

Shin-Jeng Lin
Department of Business
LeMoyne College
Linsj@mail.lemoyne.edu

Abstract

This study intends to enhance the understanding of successive searches over multiple sessions by characterizing successive searches with a conceptual model, Multiple Information Seeking Episodes (MISE), validating MISE and supporting successive searches with a prototyped information system, PERsonalized and Successive Information Seeking Toolkit (PERSIST), whose requirements are derived from MISE. The study has both theoretical and practical values as it increases the understanding of human information behavior and develops useful tools to support the concerned behavior.

Keywords: MISE, PERSIST, successive searches, conceptual model.

RESEARCH OBJECTIVE AND QUESTIONS

Information seeking for the same information problem over the web can easily be carried out in more than one session. Intuitively, a person could engage in successive searches over multiple sessions simply because of physical or mental fatigue, time constraints, etc. However, intuition is not sufficient to create a thorough understanding of users' information seeking behavior. Theoretically, from the phenomenology perspective, the knowledge state of a searcher is cumulative, continuous, and analog (Schutz and Luckmann 1973). Nevertheless, neither phenomenology nor information systems research has provided a detailed explanation about how the current search could be related to previous searches, not to mention the implications for system design. Without a theoretical foundation, we cannot explain or predict what resources needed by users in previous search sessions (e.g., query terms, pages visited) would be useful in assisting users in later sessions and how they would be useful.

To fill the vacancy of theory and bridge the gap in practice of system design, this study places its fundamental research question on how to understand problem-centered information seeking across sessions, which entail the following two sub-questions. First, how can we characterize successive searches across sessions? Second, what are the implications for system design in supporting successive searches?

The first question is explored via the creation of a conceptual model that fundamentally enhances our understanding of why and how we search for information across multiple episodes. The second question is investigated with the prototype of a Web interface that can enable the verification of the conceptual model on successive searches across sessions.

THEORETICAL FOUNDATIONS

Successive searches are an emerging research paradigm related to information science, information systems, and cognitive psychology. Several studies on this topic are currently underway (e.g., Kolomdi 2001; Spink et al. 1998; Vakkari 2001), in addition to the proposed conceptual model, **M**ultiple **I**nformation **S**eeking **E**pisodes (MISE). MISE lays its foundations on a review of the interdisciplinary literature. Lin and Belkin (2000) provide a concise version of the MISE model. A more detailed version of MISE is in Lin (2001). A brief summary of the MISE model follows.

MISE consists of the four dimensions of information seeking experience: problematic situation, information problem, information seeking process, and episode. A problematic situation is the users' subjective perception and estimate of carrying out a goal with their existing knowledge in the objective context. An information problem is the result of that perception and estimate, requiring external information to explicate. The explication of information problem is manifested as the information seeking process, constituted by the activities users engage in when interacting with information resources for information seeking. The time period in between the initiation and termination of interacting with a particular external information resource is the episode.

Each dimension of MISE has a set of associated properties. In the dimension of problematic situation, the main properties include tasks and problematic cognitive states. Tasks are the work the users carry out that requires information to explicate the problematic situation (e.g., writing a paper, planning a vacation). Problematic cognitive situations are the cognitive reasons why the users would need to acquire additional information on top of their existing knowledge in order to resolve the problematic situation.

The properties in the information problem dimension are those that would indicate the extent to which the users are capable of defining information problems and specifying the criteria for plausible treatments. For example, complexity of information problem means the number of concepts connoted in an information problem (Saracevic et al. 1988). The focus of the information problem reflects the ability to determine the primary concept in the information problem (Wang et al. 1988).

The properties in the dimension of the information seeking processes characterize how users actually interact with external information systems and internal knowledge states to search for information. Based on Marchionini (1995), three sub-processes are identified: articulation of information problem, evaluation of search outcomes, and monitoring of the problematic situation. Articulation of the information problem concerns the identification and representation of the information problem; its main properties include methods of representation (e.g., explicit queries versus implicit recognition) and clarity of representation (ambiguous versus precise). Evaluation of the search outcome determines the usefulness of the information examined; its main properties entail evaluation criteria (e.g., novelty, informativeness, credibility, etc.; see Saracevic 1996) and magnitude of criteria. Monitoring of the problematic situation examines the extent to which the information problem has been resolved; its primary properties include awareness of information attainment and awareness of evolution of the information problem.

Table 1. Renewal Reasons in Successive Searches

Renewal Reasons	Definition
Information problem is transmuted	The information problem is usually in the formation phase or gets enriched, polished, or added on with new concepts and finally changes its texture from the original state to the transmuted state.
Information problem is spawned	The information problem itself can spawn sub-problems that extend over the surface of the original problem and have a higher priority than the original problem.
Information problem is transited	Transition is a process of changing from one problem, A, to another, B. A and B are, in essence, different problems and connected with a cause-result relationship.
Information problem rolls back	The problematic situation that once was thought to have been resolved actually turns out to be infeasible or ineffective when applying treatment to the information problem.
Treatment of information problem becomes unavailable	Information objects for treating the information problem, found once, are no longer available when trying apply to them to the information problem or communicating with co-workers.
Information problem is unanswered	The information problem can be counted as unanswered if the information seeker cannot find satisfactory information objects as treatments at the termination of an information seeking episode.
Information problem is cultivated	The information problem can be cultivated in the sense that an information seeker is trying to cultivate or to stay abreast of an area of information or interest.
Information problem is anticipated	The information problem can be anticipated because some elements of the current situation match patterns of past experiences which are estimated to revive, or are appropriate for, the envisioned or planning situation in the future.

The dimension of episode mainly is concerned with reasons why users re-initiate and terminate a search session. The renewal reasons are particularly important as they signal the latest states of knowledge, cognition, and information problem of users; therefore, they imply different support needed from information systems. Eight different renewal reasons are identified and summarized in Table 1.

The properties in the lower dimensions help shape those in the upper dimensions, while those in upper dimensions could provide feedback and influence in the lower dimensions, as shown in Figure 1.
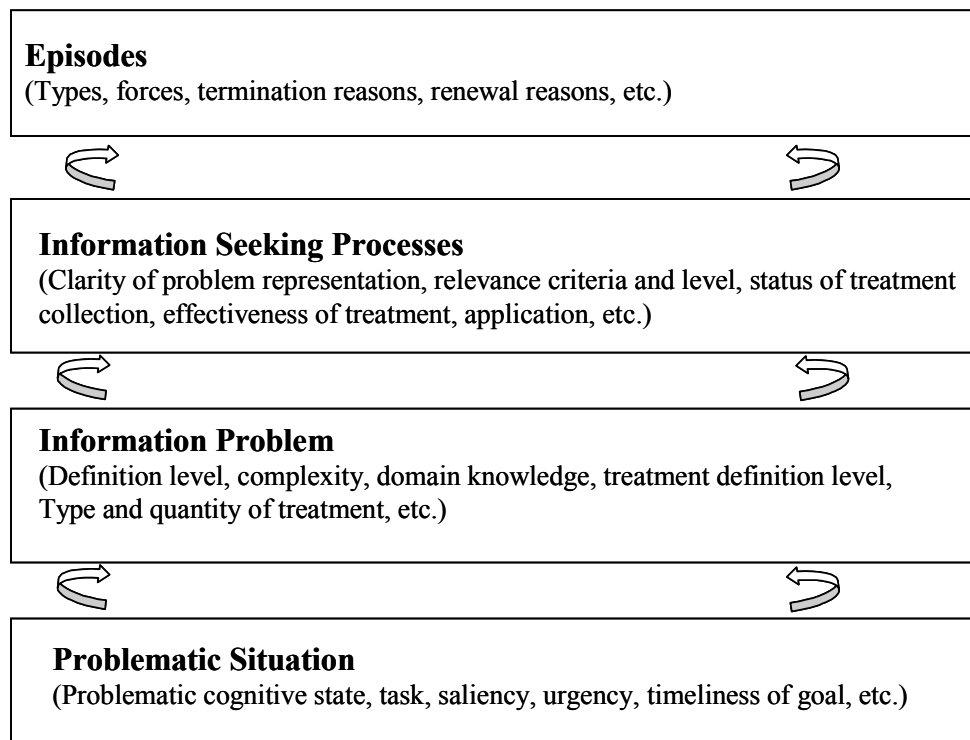

Episodes
(Types, forces, termination reasons, renewal reasons, etc.)

Information Seeking Processes
(Clarity of problem representation, relevance criteria and level, status of treatment collection, effectiveness of treatment, application, etc.)

Information Problem
(Definition level, complexity, domain knowledge, treatment definition level, Type and quantity of treatment, etc.)

Problematic Situation
(Problematic cognitive state, task, saliency, urgency, timeliness of goal, etc.)

Figure 1. Structure of MISE

RESEARCH METHODOLOGY

Overview

The study is designed in an iterative four-phase fashion. Interdisciplinary literature from information science, information systems, and cognitive psychology was reviewed to propose MISE. MISE can help not only in increasing understanding of successive searches across multiple sessions, but also in specifying requirements for prototyping an interface to support the concerned behavior, **PER**sonalized **S**uccessive **I**nformation **S**eeking **T**oolkit (PERSIST). A lab-controlled experiment in which recruited subjects used PERSIST was carried out to explore the answers of the two research questions. The results of the experiment helped in the revision of MISE, the adjustment of system requirements, and in devising new support functions.

For research question one, if the experiment finds all of the key properties of MISE in subjects' successive search experiences, the sufficiency and accuracy of MISE in characterizing successive searches are verified. For research question two, if the experiment finds that PERSIST is effective in reducing the information seeking problems identified with MISE, then the predictive power of MISE is verified. The variables in exploring effectiveness of PERSIST are those adopted from HCI and information retrieval studies. For example, the usability measures are based on a series of interface studies on information retrieval systems (Belkin et al. 2001).

Experiments

Task

Due to the complexity of successive searches over sessions, the study could only focus on a limited set of renewal reasons at a time, electing to focus in particular on the transmuting mode of successive searches, where the users gradually develop the foci of their information problems. A concrete scenario of the transmuting successive searches task is when the users attempt to search for information to plan a vacation when they do not have exact destinations and have limited knowledge about prospective destinations. Subjects were given the scenario and required to engage in three experimental sessions with one to three days in between sessions. Subjects could stop any time after 20 minutes, but had to stop at 30 minutes. The flexibility in termination time is important as it provides the opportunity to observe the termination reasons.

Subjects

Twenty subjects were recruited from an undergraduate class on communication research methods, with equal numbers of male and female students.

Data Collection

Subjects were required to participate in interview questionnaires before and after each search sessions. The interview questionnaires were semi-structured and designed to reflect constructs that would help answer the two research questions, as shown in Table 2. An online log was used to record the usage of PERSIST.

Data Analysis

For research question one, content analysis is utilized to identify the key variables (e.g., those under the "interaction" category, such as termination status and reasons) in subjects' successive search experiences based on the results of the interviews. For research question two, the paired t-test helps examine whether there are significant differences in key variables of the information seeking process between sessions (e.g., abilities in articulation and evaluation, level of knowledge). Pearson correlation helps explore whether usability and usage frequency of PERSIST could explain why subjects perceived PERSIST as having different levels of effectiveness. Additional qualitative interview data (e.g., motivation, understanding) would help triangulate subjects' actual intents.

SUMMARY

Contributions of the Study

The significance of the study is multi-dimensional. First, it enriches the literature in understanding successive search behavior with the MISE model. Second, the MISE model can serve as the foundation of task analysis for software engineering. Third, a prototyped interface based on the design implications derived from the MISE model can actually yield useful features in supporting successive searches.

Current Status of the Project

The status of the study is currently at the fourth phase. This phase entails observation of subjects who use PERSIST to search for vacation information on the Web in a controlled lab environment. The experiment has been conducted and all data collected. Currently, the experimental results are being analyzed. The revised MISE model will be available for conference presentation.

Table 2. Constructs Observed in the Experiment

Measures	Conceptual Definition
Motivation	Reasons for using PERSIST features
Understanding	Understanding of purposes of PERSIST features
Usage Frequency	Number of times of using PERSIST features
Effectiveness:	Users' subjective satisfaction about the *experience* and *consequence* of the interaction with PERSIST to gratify their renewal motivation
Seamlessness	The extent to which PERSIST allows users to "smoothly continue" successive search episodes
Productivity	Usefulness of information attainment
Performance	Performance of information seeking
Ease in Articulation	The extent to which PERSIST helps representing users' information problems
Ease in Evaluation	The extent to which PERSIST helps assessing relevance of search results
Ease in Monitoring	The extent to which PERSIST helps monitoring the progress of accomplishing the task
Usability:	System's capability in letting the users do whatever they intend to do at the first place
Understanding	Ease of understanding what PERSIST features are intended for
Learnability	Intuitive usage of PERSIST features
Intuitive usage	Ease of actually using PERSIST features
Affinity	The extent to which users like to use PERSIST features
Adoption likelihood	Possibility of using PERSIST features for other similar search tasks
Interaction:	
Goal	Reasons users search for information
Level of knowledge	Amount of knowledge users possess or gain
Definition of information problem	To what extent users have defined their information problems
Definition of treatment	To what extent users have defined their treatments
Relevance criteria used	Criteria used to determine information objects that helps cope with information problem
Strength of relevance criteria	Level of relevance criteria in determining information objects as treatment
Information assimilation	Amount of information users learn from information objects they interact with
Ability in articulation	Ability to articulate information problems
Ability in evaluation	Ability to evaluate search results for treatment
Ability in monitoring	Ability to self-estimate the progress toward accomplishing the ultimate goal
Termination status	Termination status of an information seeking episode
Termination reason	Termination reason of an information seeking episode

References

Belkin, N., Perez-Carballo, J., Cool, C., Kelly, D. J., Lin, S., Park, S. Y., and Sikora, C. "Iterative Exploration, Design and Evaluation of Support for Query Reformulation in Interactive Information Retrieval," *Information Processing and Management* (37:3), 2001, pp. 403-434.

Komlodi, A. *Search Histories for User Supporting in Information Seeking Interfaces*, Unpublished Ph.D. Dissertation, University of Maryland, College Park, 2001.

Lin, S. J. *Modeling and Supporting Multiple Information Seeking Episodes*, Unpublished Ph.D. Dissertation, Rutgers University, New Brunswick, NJ, 2001.

Lin, S. J., and Belkin, N. J. "Modeling Multiple Information Seeking Episodes," in *Proceedings of the 63rd ASIS Annual Meeting,* Chicago, IL, 2000, pp. 133-147.

Marchionini, G. *Information Seeking in Electronic Environments,* Cambridge University Press, New York, 1995.

Saracevic, T. "Relevance Reconsidered," in *Proceedings of Second International Conference on Conceptions of Library and Information Science (COLIS),* Copenhagen, Denmark, 1996.

Saracevic, T., Kantor, P., Chamis, A. Y., and Trivison, D. "A Study of Information Seeking and Retrieving: II. Users, Questions, and Effectiveness," *Journal of the American Society for Information Science* (39:3), 1988, pp. 177-196.

Schutz, A., and Luckmann, T. *The Structures of the Life-World* (R. M. Zaner and H. T. Engelhadt, Jr. trans.), Northwestern University Press, Evanston, IL, 1973.

Spink, A., Wilson, T., Ellis, D., and Ford, N. "Modeling User's Successive Searches in Digital Environments," *D-Lib Magazine*, April, 1998 (http://www.dlib.org/dlib/april98/04spink.html).

Vakkari, P. "A Theory of the Task-based Information Retrieval Process," *Journal of Documentation* (57:1), 2001, pp. 44-60.

Wang, X., Liebscher, P., and Marchionini, G. "Improving Information Seeking Performance in Hypertext: Roles of Display Format and Search Strategy," Technical Report CAR-TR-353, Center for Automation Research, University of Maryland, 1988.