

Types of Document Search Tasks and Users' Cognitive Information Seeking Strategies

Hee-Eun Lee¹ and Wan Chul Yoon²

¹ Department of Industrial and System Engineering, KAIST
291 Daehak-ro, Yuseong-gu, Daejeon, 305-701, Korea

² Department of Knowledge Service Engineering, KAIST
291 Daehak-ro, Yuseong-gu, Daejeon, 305-701, Korea
{lhe1008, wcyoon}@kaist.ac.kr

Abstract. For the researchers and learners, an unprecedented number of documents became available on the Internet and academic archives. Powerful search systems and sophisticated recommendation services are also available. Despite the IT assistance, finding the most useful information in daily knowledge works has become a cognitively demanding task more than ever due to the overwhelming number of documents. To improve the search systems with better human-computer cooperation, human information seeking strategies should be understood. This paper reports a study that identified the differences in the user search strategies with respect to two major search task types: open and purpose-driven exploring (OT) vs. closed and target-specified (CT) tasks. An observational experiment was conducted and the results were analyzed by mapping the user activities on a cognitive task-flow framework. The analysis comparing user activities in four search tasks revealed notable differences in their strategies to deal with the two task types. More frequent re-planning, especially goal reformulation, was observed for OT type tasks. The difference indicates that OT type tasks tended to trigger more knowledge-based behavior, while CT type tasks were performed relying more on rule-based behavior. These findings provide important insights for the design of search systems and user interfaces of knowledge-based systems.

Keywords: Information Search, Information Seeking Strategies, Task Types, Interaction Design, Decision Behavior.

1 Introduction

With the development of various smart-devices, such as smart-phones and tablets, information search activities have become daily tasks. With the rapidly increasing amount of content available in the Internet, people became more relying on the knowledge on the Web for learning, entertainment, research, shopping, and other daily activities. Search is in the center of the user activities interacting with the Web. Some researchers pointed out that the development of Web search systems has not sufficiently reflected the current situations of the increased Web usage and the growth of Web contents [3]. To overcome the limitations of the current systems, researchers have tried to develop various Web technologies including Semantic Web technologies

and natural language search technologies [1, 9]. Despite the development of diverse innovative technologies, however, most Internet users still choose to stay with general Web search sites, such as Google, that are based on keyword search method. It might be noted that the most previous researches have focused on experimenting with Web search system design, not on investigating the users' Web search behaviors and epistemic user models. Some researchers have asserted that actual users' Web search tasks should be analyzed and applied in the process of Web system design to resolve the problem [4-5].

Web search is one of the keys to the Web's success [12]. In the academic circle, the research subjects of Web search are classified into three terms according to the targets of the search process, which are information retrieval (IR), information behavior (IB), and information seeking (IS) [4]. Regarding these classification, Järvelin [5] has described their differences in his paradigm model that IR focuses on the Web system, IB focuses on users' searching activities, while IS refers to the interaction between users and Web systems. Similarly, Xie has discussed the differences between IR and IS [13]. Based on his definition, IR is similar to information seeking (IS), but is more limited to the use of computer systems. In contrast, IS refers to purposive behavior including users' ways of using IR systems in order to pursue their information goals [13]. In this study, we considered Web search activities as the process of IS from the viewpoint of human decision making strategies.

Xie mentioned that research on information searching might be focused on different levels, which are tactics/moves, strategies, usage patterns, and models. His planned-situational interactive IR model assumes that the user would start from high-level goals or tasks to establish a plan to achieve them. The plan is then realized through information seeking strategies that comprises interactive intentions and retrieval tactics. After interaction with the IR interface, the user considers the resulting situation to decide whether the cycle should be continued according to the plan or disrupted to modify/replace the plan relying on his/her knowledge.

There have been some previous works regarding the effects of factors influencing information search strategies. Marchionini studied the different tendencies of search strategies in relation to open tasks and closed tasks. He designed an exploratory study for elementary school children to search an electronic encyclopedia on CD-ROM and found differences in information seeking strategies (ISS) according to tasks [7]. Navarro-Prieto conducted a study to determine the characteristics of ISS in accordance with the experience levels of searchers in given tasks. He classified the participants as experts and novices on specific search tasks. As a result, he could identify top-down and bottom-up strategies depending on the experience levels [6]. Some other researchers have been interested in differences in ISS according to age. Aideen studied the degree of searching capability and the difference in ISS pursuant to users' ages [10]. Similarly, Chin carried out a study on the effects of users' age that determine the types of ISS for medical diagnosis searching tasks [2]. The results of their works showed that older users tended to depend more on browsing strategies while younger users used more active exploring strategies.

Rasmussen's works [8, 11] in the field of cognitive engineering help understand how the ISS is formed and how the task types influence users' cognitive ISS. Rasmussen conducted research to understand human diagnostic strategies to locate failure points in machines. He carried out strategy analysis that identified four major types of fault-search strategies and represented them in information flow diagrams. In another study, based on cognitive task analysis for nuclear power plant operation, he established a framework that explains task-performing strategies, which was referred to as the decision ladder [8, 11]. For some tasks, the decision ladder can be fully traversed through the stages of acquisition of observable data, interpreting them to identify the current state, evaluating the state taking the system goal into account, determining necessary tasks, forming an appropriate plan, and finally to the implementation of the plan. However, depending on the characteristics of tasks and situations as well as human task knowledge, the process can partly be omitted using various shortcuts. The full decision making path tends to appear in *knowledge-based decision behavior*, while in *rule-based decision behavior* the high-level decision stages (e.g., goal consideration or task formation) are largely omitted due to accumulated experience.

Similarly, in the process of information seeking, it is reasonable to expect that task types together with the user's knowledge on the subject will shape the user's strategy. It is very important to find out the relationships between the task types and search strategies for developing effective search systems or devising suitable supporting features. With this motivation, we examined two types or modes, of information seeking tasks in this study; one is more open, exploring-like search and the other is more closed, tightly specified search. Those modes are expected to affect the degree of reliance on the knowledge-based or rule-based behavior in search.

We constructed an *IS task-flow framework*, a task-specific variation of Rasmussen's decision ladder, to analyze and explain the observed user activities and find evidences for the users' search strategies. The framework is in part similar to Xie's planned-situational interactive IR model, but it is designed to account for ad-hoc browsing tactics besides plan-initiated search actions. Using the framework, this study aims to identify users' actual searching strategies and find the effects of task types on the information-seeking strategies.

2 Hypotheses

While understanding human information seeking strategies is essential for designing better search systems and user interfaces, it may be pointless to establish a single model of human information seeking considering the diversity of task types that may shape the human strategies. There may be no universally optimal user interface design that facilitates all different types of search tasks.

We define two contrasting categories of search tasks to investigate the variation in human strategies. One is called OT (Open Task) type and indicates a search mode in which the searcher is relatively freer in pursuing target information. The search is more like exploration in knowledge space although the user has a solid information

seeking purpose, not merely surfing or wondering. The task might be performed within a theme or a topic and have some criteria for usefulness of the finally acquired information. However, besides the purpose and quality to pursue, there may be little strict specification of target documents in terms of attributes or containing contents. An example of OT type task is “We need to prepare for a presentation on the topic of human memory. Get acquainted with the subjects and collect useful set of contents” Then, the user should find and choose relevant subtopics, theories, and cases that would be useful for the presentation. The user is free to select and organize information except that the information should effectively serve the given purpose.

The other search task type is CT (Closed Task) in which the user works with more narrowly specified attributes of the target to find. An example is “Regarding semantic web, what are its definition and important technical ingredients?” CT type search is undertaken when the user has to find some documents that contain specific contents.

In this study, we are mainly interested in information search under the purpose of learning. The classification of OT vs. CT is practically meaningful in learning situation. Both types of tasks are frequently experienced by students in searching learning materials on the Web and may lead to an important distinction in their strategic behavior during information seeking. OT type search typically appears in the early stage of learning new subjects or trying to grasp overall terrain of the domain knowledge of interest.

The criteria of performance may also differentiate between the two task types. In CT type search, fitness of the acquired information gets importance while in OT type the quality of information is emphasized. In CT type, more attention may be exercised to determine whether a particular document contains the information being looked for or falls in the range of target documents. In OT search, the user may be more engaged in evaluating various documents and accordingly adjusting the direction of search. These differences could closely be related with the distinction between knowledge-based and rule-based behavior.

We did not hypothesize on very detailed strategic elements that are expected to appear in the two types of tasks. The experiment was conducted largely as a primary observation to probe any patterns of activities that can separate a type of tasks from the other. Therefore, the hypotheses are defined at a rather high level, stating that

- ◇ There will be strategy differences between OT and CT type search tasks that can be found in terms of decision behaviors.
- ◇ The differences may be related with more generalizable distinction between knowledge-based behavior and rule-based behavior.

3 Experiment

To understand the effects of task types on information seeking strategies, an experiment was conducted with four searching task scenarios on a document search system. In the experiment, twenty participants carried out search tasks to achieve pre-specified goals and their actual searching actions were recorded and analyzed using the IS task-flow framework.

3.1 The System and Participants

The System. *CourseShare* search system (Figure 1) was developed at KAIST to support its students and educators to efficiently find the most suitable educational documents among overwhelmingly abundant materials available on the Web. The system is still in the experimental stage and growing in the number of indexed documents. We used about 1.1 million slide-type documents in this experiment. The search tasks were performed with the support of *CourseShare*'s various search functions including keyword-based content search and attribute search with such slots as institute, author, and date, etc. The search system had a Web-based interface that was composed of main page, search list page, and detailed information page.

The screenshot displays the CourseShare search results page for the query 'information systems'. The page features a search bar at the top with the text 'information systems' and a 'Search' button. Below the search bar, there are navigation links: '<< Back to the Result' and 'Advanced'. The main content area shows the search results for 'IMS3230 - Information Systems Development Practices'. The results include the course title, lecturer information (Andrew Barnden, Monash University), a brief description, and a list of 15 related documents. Each document entry includes a title, a brief description, and a rating. On the right side, there is a sidebar with a 'Table of Contents' and a 'Knowledge Network' section. The page also includes social media sharing options (Twitter, Facebook) and a 'Rating Now' button.

Fig. 1. The CourseShare Search System for Learning

Participants. Twenty students, 4 undergraduate and 16 graduate students (14 males and 6 females) participated in the experiment. The average age was 26 years with a standard deviation of 2.92. They were recruited from diverse majors to exclude particular training effects of a specific department. All participants were familiar with Web search tasks. Thirteen subjects answered that they usually performed Web search to find learning materials more than one time per day during their study. The other six students usually searched information on the Web for learning purpose two or three times a week. Only one student answered that he tended to search learning materials once in a week. However, all the subjects used *CourseShare* system for the first time in this experiment. Therefore, no previous learning of the given experimental system could influence the strategies and performance of the participants. Before the

experiment session, participants were given simple five-minute training tasks, when the search system's characteristics and available features were instructed.

3.2 Tasks

Four task scenarios, two for each task type respectively, were devised with different themes or subjects. The themes of task scenarios and their task types are shown in Table 1.

Table 1. Task types and theme of task scenarios for the experiment

<i>Task #</i>	<i>Theme of task</i>	<i>Task type</i>
1	Human memory/ Memory theory	OT type
2	Media	
3	Semantic web/ Semantic web service	CT type
4	Information systems	

Each participant was given all of the four search scenarios in random orders, and a brief post-experiment survey was carried out. No explanation of the designed task types was given to the participants to ensure the accuracy of the test. A task ended after the participants downloaded a few most useful documents from the system. The average length of the search time for a task was around 10 minutes although the duration was not forcibly controlled because the purpose of the experiment was identifying the characteristics in search strategies, not evaluating the search performance. To provide a non-distracting task environment, the experiment was conducted with one or two participants at a time. The whole processes of information search were recorded employing a screen capture recording program, Free Screen to Video version 1.2 (Koyote software, Cyprus). In addition, the log data of users' search tasks were collected for analysis.

3.3 Framework-Based Analysis

The recorded video data and log data that were collected during the experiment were analyzed to investigate the users' information seeking activity and identify characteristic strategies employed by different groups and in different tasks. Data analysis was performed to identify the users' actions in terms of information seeking task flow. The results were later used for the evaluation of task categorization, which was to check the significance of the hypothesized task types as well as the fitness of the given tasks to the categories. Finally, strategy analysis vs. conducted to comprehend the information seeking process and find the effects of the two task types, or modes, on the information search behavior.

IS Task-Flow Framework. Human information-seeking activities can be represented as transitions in our *IS task-flow framework* that is shown in Figure 1. The whole search process was represented as a dynamic model aggregating eight search actions

(i.e., the rectangles in figure 2), namely goal formulation, query formulation, search, system tool selection, browsing, selection and evaluation, utilization, and changing plan. Among them, four points of decision-making are of our special interest in association with information seeking strategies: query formulation (QF), browsing (BR), selection/evaluation (SE), and changing plan (CP). At the point of QF (query formulation), the decision is to choose a search keyword considering the search goal and the search history to the point. Next, at the point of BR (browsing), the user visually reviews the searched list to evaluate the results and find one or more interesting materials. SE (selection/evaluation) is a combination of a few processes: selecting an interesting document, opening the document, and evaluating whether the selected document is satisfactory or not. For example, if a user is not satisfied regarding the contents of a selected document, then she or he may click the back button and visit the list page (BR process) again to find other materials without changing the search plan. Otherwise, the user may move to the point of CP (changing plan) to re-plan the search. The re-planning may be changing the search direction and/or formulating another search keyword as they are expected to result in better searching.

All actions that the participants performed during the search task process were identified and put down on the corresponding transition flows (i.e., arrows) on the IS task-flow framework. The frequency of each transitional flow was accumulated. The relative frequencies of the transitions were calculated taking the total number of action steps as the denominator. Additionally, the number of search keywords and the performance time were checked in the data analysis.

4 Results

The statistical results of the experiment show that the task scenarios were properly devised to represent the two task types. Also significant differences are found in the information seeking process between the task types.

4.1 Task Performance

The search task activities of the 20 participants are summarized in Table 2. As seen in the results, OT types showed higher numbers of total actions during the search process than CT types. The total number of actions is the same as the sum of actions mapped onto the IS task-flow framework. OT type tasks also require more number of search keywords than CT types. In terms of the average task time for a search task, participants generally took longer time to perform OT types than CT types although the difference was not as big as the number of actions or keywords.

The results shown in Table 2 indicate that the influence of task type on the search process is notable on the whole. More numbers of action steps and keywords were needed for the OT type tasks as might well be expected. This result also confirmed that the categorization of OT vs. CT was meaningful and the four task scenarios appropriately represented the two types of tasks. More detailed analysis on how the task types affected the users' cognitive behavior in information seeking is called for. It is discussed in the next section.

Table 2. The results of task performance (twenty participants on four search tasks)

<i>Task</i>	<i>Task type</i>	<i>Total number of action steps</i>	<i>Total number of search keywords</i>	<i>Average task time</i>
1	OT type	1,027	99	11' 41'' 33
2		1,233	129	11' 40'' 39
3	CT type	736	66	9' 54'' 51
4		682	47	10' 57'' 30

4.2 Information Seeking Strategies

Using the IS task-flow framework, we analyzed the participants' search processes. Then we compared the effects of task types (OT types and CT types) on the strategies, by considering the recorded actions at the points of decision-making during the search process.

Re-planning Behavior. The results of description of the task-flow framework are shown in figure 2. As seen in the figure, differences are apparent between task types in the search process. For OT types, the results show that searchers change their search plans much more often during search tasks. Thus, for OT type tasks, search goals were more often reformulated and search keywords were also refreshed more times, whereas for CT types users tended to repeat the browsing process longer without changing the search plan. It is also noted that when changing search plan (CP), the relative rate of goal reformulation over mere change of queries is more than 8 times in OT, while the ratio stays within 3 times in CT type.

Decision Frequencies. It was mentioned that, in the task-flow framework, four points of decision-making are of our special interest since they may help identify information seeking strategies: query formulation (QF), browsing (BR), selection/evaluation (SE), and changing plan (CP). We performed qui-square tests between OT and CT types for the frequencies those decision points were visited during search. The results are summarized in table 3. At all four points of decision-making, significant differences are found between task types. Users show a tendency to perform query formulation (QF) and changing plan (CP) more in OT type search tasks than CT type. In contrast, browsing (BR) and selection/evaluation (SE) was more often performed in CT type search than OT type.

Table 3. The analysis results of the effects of task types

<i>Task types</i>	<i>QF</i>	<i>BR</i>	<i>SE</i>	<i>CP</i>
OT vs. CT (Task 1,2 vs. Task 3,4)	Difference (p=0.046)	Difference (p=0.006)	Difference (0.030)	Difference (p<0.0001)
Comparison	OT > CT	OT < CT	OT < CT	OT > CT

missions. In two missions, the goals of the search were well-specified with clearly stated usage, while for others the goals were somewhat loosely defined, e.g. to collect a set of useful documents. To study the difference of user behavior, a superset model of search activities was established the observed transitions between elementary activities were recorded and compared.

Regarding the decision frequencies, participants changed the search plan, especially reformulating the search goal, more often in OT type tasks than CT type search. On the other hand, for CT types, participants' actions were more focused on the process of browsing and selecting/evaluating without alteration of search goal or search keywords.

The IS task-flow framework was devised tailoring Rasmussen's decision ladder for search tasks. The above mentioned differences provide good evidence that the OT type tasks require more knowledge-based decision making while CT tasks rely more on rule-based decision making. This has very strong implications in system design and interaction design because the characteristics of knowledge-based and rule-based decision-making behaviors have long been studied and are well understood.

The different search strategies in the two task types found in this study provide important insights for designing the search interaction, user interface, and support functions. Reformulating search plans and repetitive browsing/evaluating the searched information are two most important and very different subtasks. According to the task types, the more emphasized subtasks should intensively be supported by interface design or supporting features.

In an early stage of approaching a new subject, people should make rather abstract decisions such as interpreting the information in currently found documents, comparing or integrating the information from various sources, and evaluating the information considering the overall purpose of search. This in turn will suggest new directions of search or new set of keywords to use. New information acquired by the adjustment of plans or keywords are again evaluated to decide the direction was effective and should be continued or should be given up. It means that, to perform tasks at the level of knowledge-based behavior, much cognitive effort is required for users [8]. To help, the history of the search session, a list of related topics and key concepts, search by good examples, and an intelligent interface that supports the full cycles of such conceptual exploration may be very effective for OT type search tasks. Also knowledge-based reasoning depends on general background knowledge, rather than the search tactics, of the users. Therefore taking users' knowledge structure into account in interface or aiding, if at all possible, may also greatly help enhance human-system cooperation in OT type search.

For CT type tasks, users tend to perform the information search relying heavily on browsing with less frequent changes of search plans. Since the targets of search are well specified, the user concentrates on evaluating the fitness of found documents to the target. A rapid decision-making regarding the closeness of the documents to the set target is required. The search system could assist the user in such CT type tasks with more accurate recommendation features, filtering with easily adjustable and combinable criteria, deeper text mining functions, and some measures or guiding

visualization that provides feedback on how the search is successfully approaching toward the target.

It should be noted that, while OT type tasks show similar characteristics with knowledge-based decision problems, it does not necessarily mean that OT implies that the user's training is low in the subject area. Neither does it mean the user can be trained to use CT type search instead of OT for the same problem. The knowledge-based vs. Rule-based dichotomy is about the decision behaviors and in many cases depending on the experience level of the user. OT-CT dichotomy is about the characteristics of the search task itself that tend to force one of the decision modes regardless of the human knowledge or training. In OT type tasks, the search decisions require more high-level or abstract knowledge, hence are more difficult for novices in the domain knowledge to handle.

Future work will be directed toward further clarification of the relationships between the user's background knowledge and the search tasks. As both the importance of search tasks and the possible technologies to assist knowledge search are increasing, studies on human information seeking strategies will become increasingly more valuable as the foundation of search system improvement.

Acknowledgement. This research was supported by Korea National Research Fund from 2011 to 2014.

References

1. Antoniou, G.: A semantic web primer. The MIT Press (2004)
2. Chin, J., Fu, W.T.: Interactive effects of age and interface differences on search strategies and performance. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 403–412. ACM (2010)
3. Hu, W.C., Chen, Y., Schmalz, M.S., Ritter, G.X.: An overview of world wide web search technologies. In: Proceedings of 5th World Multi Conference on Systems, Cybernetics, Informatics, SCI 2001, Orlando, Florida, pp. 22–25 (2001)
4. Ingwersen, P., Järvelin, K.: The turn: Integration of information seeking and retrieval in context, vol. 18. Springer (2005)
5. Järvelin, K., Ingwersen, P.: User-oriented and cognitive models of information retrieval. Understanding Information Retrieval Systems: Management, Types, and Standards 47 (2012)
6. Navarro-Prieto, R., Scaife, M., Rogers, Y.: Cognitive strategies in web searching. In: Proceedings of the 5th Conference on Human Factors & the Web, pp. 43–56 (1999)
7. Marchionini, G.: Information-seeking strategies of novices using a full-text electronic encyclopedia. *Journal of the American Society for Information Science* 40(1), 54–66 (1989)
8. Rasmussen, J.: *Information Processing and Human-Machine Interaction. An Approach to Cognitive Engineering* (1986)
9. Sánchez, D., Moreno, A.: Development of new techniques to improve web search. *System* 4(4), 2 (2005)

10. Stronge, A.J., Rogers, W.A., Fisk, A.D.: Web-based information search and retrieval: Effects of strategy use and age on search success. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 48(3), 434–446 (2006)
11. Vicente, K.J.: *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. CRC Press (1999)
12. Wilson, M.L., Kules, B., Shneiderman, B.: From keyword search to exploration: Designing future search interfaces for the Web. *Foundations and Trends in Web Science* 2(1), 1–97 (2010)
13. Xie, I., Bates, M.: Information searching and search models. *Understanding information retrieval systems*, 31–46 (2012)