SketchBrain: An Interactive Information Seeking Interface for Exploratory Search

Hogun Park, Sung Hyon Myaeng, Gwan Jang, Jong-wook Choi, Sooran Jo, Hyung-chul Roh
School of Engineering, Information & Communications University (ICU)
119, Munjiro, Yuseong-gu, Daejeon, 305-732, Korea
+42-866-6210, 82
{gsgphg, myaeng, ilys23, pudidic, ddangly-, nuunmir }@icu.ac.kr

ABSTRACT
As the Web has become a commodity, it is used for a variety of purposes and tasks that may require a great deal of cognitive efforts. However, most search engines developed for the Web provide users with only searching and browsing capabilities, leaving all the burdens of manipulating information objects to the users. In this paper, we focus on an exploratory search task and propose an underlying framework for human-Web interactions. Based on the framework, we designed and implemented a new information seeking interface that helps users to relieve cognitive burden. The new human-Web interface provides a personal workspace that can be created and manipulated cooperatively with the system, which helps the user conceptualize his information seeking tasks and record their trails for future uses. This interaction tool has been tested for its efficacy as an aid for exploratory search.

Categories and Subject Descriptors
H.3.7 [Information Storage and Retrieval]: Information seeking Interface

General Terms
Documentation, Design, Experimentation.

Keywords
Information Seeking Interface, Exploratory Search

1. INTRODUCTION
For a traditional Web search engine, the process of querying and viewing the results is usually regarded as a single, isolated session that ends in itself. As the Web has become a commodity, however, it is used for a variety of tasks in many different ways, encouraging new paradigms in information seeking (e.g. berrypicking [1], information foraging [2], and sense-making [3]). However, most popular commercial search engines have taken a conservative position and adhered to the traditional model, leaving all the rest of the information seeking and related tasks to the user. More specifically, the user has all the burdens of manipulating the information objects that have come to his attention in a series of search activities.

HCIR'08, Oct 23, 2008, Redmond, WA.

An area in which this type of cognitive burden affects significantly is exploratory search. An exploratory search task [4][5] is to investigate on the background information of a topic or gather information sufficient to make an informed decision. For example, assume that a user is considering purchasing a DMB (digital multimedia broadcasting) receiver. The user would want to learn more about the DMB technology and the manufacturers of various products related to it, so that he can select the provider and the products that best suit the needs. We believe that most existing search engines and their interfaces are not satisfactory for exploratory tasks, because of the following.

First, compared to the task of searching for specific or known items, an exploratory search task usually requires users to send a series of queries during a search session, visit more new domains, and revisit previously visited sites (especially branch pages) [5]. These activities together mean a significant amount of information and workload that traditional search engines have rarely attempted to reduce. The workload is associated with representing information needs [14], determining informativeness [15], and memorizing previously explored information [16]. Without explicit support from a search engine, the difficulties resulting from the workload are left as a cognitive burden to the user. Second, there are narrow interaction channels for incorporating user interests. In an exploratory search, a user needs to build up background information on a topic gradually until she feels that a sufficient amount of information has been gathered for the given task. As such, it is important to incorporate the users' interest and the information that has been found as the system processes the current query. However, current search systems rarely support the notion of "session" and interactions explicitly. While the one-time query/result model is simple and natural with HTTP, it ignores what has been done by the user in her attempt to change her anomalous state of knowledge [17]. Although there have been some attempts to infer user interest explicitly [7][8][9], implicitly [18], or both [19], the problem remains challenging, especially within the context of user-system interactions.

Given the limitations of traditional search engines for an open-ended, exploratory search task, we propose a new interaction tool that can provide an interface between a user and a search engine, called *sketchBrain*. Our aim is to provide an effective interaction environment that facilitates the series of activities in an exploratory search of the Web.

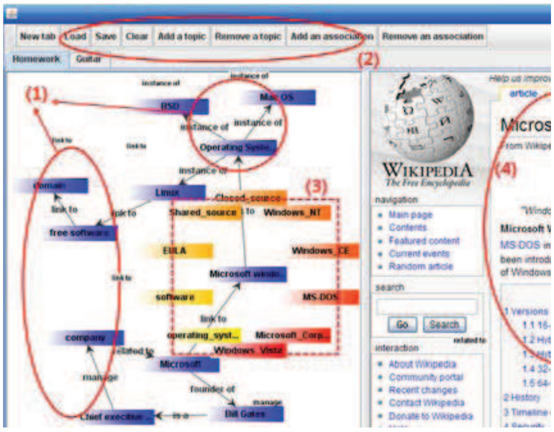

Figure 1. An example screen shot of *sketchBrain*

There are several noble features in this interaction environment. First of all, *sketchBrain* keeps track of query trails and post-query navigation trails (based on the click stream following the issued queries) and allows the users to conceptualize them. For an information seeking activity, a trail is sketched on the user's workspace of *sketchBrain*. Over the trail, the user can associate user-defined topics and system-provided semantic associations between topics using the annotation facility in *sketchBrain*. The annotation over trails means cognitive structure or the explication of user's conceptual view of the information objects being explored through interactions with the Web. It represents users' information need and affects next cognitive behaviors, so it plays an important role of reducing cognitive burden. Moreover, it has a potential for making personal metadata that can be shared with others and improving searching/browsing capability. In essence, the workspace serves as a rich memory for the past and current search efforts, which can be accessed later.

Second, our interaction tool is equipped with operations on the objects created and manipulated in the workspace. In addition to the annotation facility, *sketchBrain* allows users to manipulate the objects for their information seeking tasks. Implicit operations such as project, select, and classification (to be described in Section 3) can be utilized for the activities necessary for an exploratory search.

Third, *sketchBrain* has an intelligent path recommendation algorithm that can help users choose the most promising page to be explored at the next step in navigation. It assists users in determining informativeness of the pages that can be explored at the next step quickly.

A screenshot containing the user interface of *sketchBrain* is shown in Fig. 1. On the left is the user workspace where three workflows are sketched as indicated by (1). Using this tool, click-through data can be recorded as much as the user wishes to remember for future use. For example, whenever a user visits a new page, a new node is created and connected to an originated page or a query with a directed edge. They can be modified by manipulation tools (2), and, via this manipulation and the workspace, the user represents own conceptual understanding. In addition to this feature, our system can provide the relevant context of a specific page (like the one pointed by (4)) through time-variant multiple spreading activations (3), which can be used as a guidance for further navigation. The degree of relevance is determined by the algorithm and is shown in various colours (red indicates the most relevant one).

The remainder of this paper is composed of underlying model (Section 3), the interaction framework for supporting an exploratory search task (Section 4), and empirical evaluation via user studies (Section 5.)

2. RELATED WORK

Various information seeking interfaces have been proposed to support complex information seeking activities. Sketchtrieve [6] employs Cognitive Dimension Framework to map out the design space and provides an unstructured canvas. In this canvas, searchers can freely represent queries and corresponding search results with an intuitive interface by using typographic and layout cues that lie outside of a formal notation. Buchanan et al. [7] introduces information seeking workspace called Garnet. They exploit implicit knowledge that can be discovered from the contents in the workspace and try to find direct connections between the workspace and digital libraries. They utilize spatial parsing to extract profiles of documents and use them to learn a lexical classifier. This classifier is to identify newly searched documents that are relevant to each parsed cluster. Martin and Jose [8] suggest a personal information retrieval tool that employs a folder-like structure, so that searchers can bundle search results into folders. In addition to the interface that searchers can freely organize results, it assists query formulation and recommends hot relevant documents to each folder. Harper and Kelly [9] employ a topical structure for relevance feedback. Their interface allows users to save documents in user-defined piles for similar documents, which could be used for relevance feedback. These approaches suggest new information seeking environments with some assistance. However, their design goals are not to support exploratory search explicitly, and the systems were not tested as such. Our interface provides users with a cooperative workspace and a proactive assistance, explicitly aiming at exploratory searching tasks.

3. THE UNDERLYING MODEL

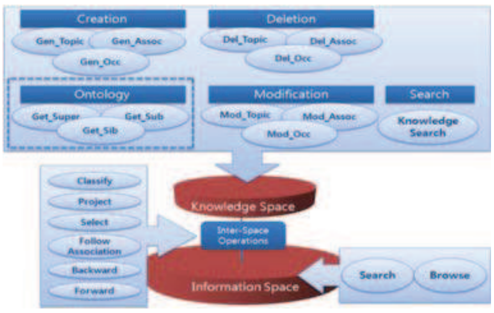

Figure 2. A conceptual view of the two-level model

Our interaction tool and user interface are based on our two-level model that explicates information and knowledge spaces where user information seeking activities take place. Fig. 2 depicts a conceptual view of the underlying model and the relationship between the information and knowledge spaces and the operations.

We attempt to separate users' conceptual work space into two levels and define operations on each space and inter-space operations [see [10] for details]. The set of operations in Fig. 2 is by no means complete, and we intend to expand it as additional needs arise.

4. INTERACTION FRAMEWORK

We have designed an interaction framework and implemented a prototype system, called *sketchBrain* that includes a search engine and the interaction tool, capturing the key ideas of the two-level model described before. *sketchBrain* is implemented with an open source graphics library (http://www.jgraph.com/) in Java, which we extended for our purposes.

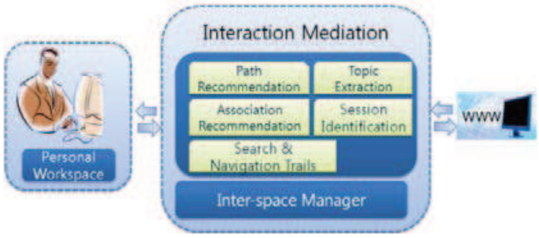

Figure 3. *sketchBrain* interaction framework

As in Fig. 3, the framework connects users with the Web through Interaction Mediation. While a user has a virtual workspace, the Web side is assumed to have a conventional search engine and browsing facilities. When the user searches/navigates the Web and attempts to make informed decisions based on the information found, Interaction Mediation provides a support with the goal of relieving his cognitive burden in the information seeking process. It consists of various tools that facilitate users' information seeking activities in terms of searching and browsing and work space creation/manipulation. Modules for topic extraction, association recommendation, and searching&browsing trails tracking assist cooperatively to construct personal knowledge structure on workspace. Path recommendation and session identification help to facilitate interaction between users and digital library. Inter-space Manager associates the personal cognitive structure with raw information in WWW and provides facilities to manipulate them. For further motivation and details of *sketchBrain*, please refer to [11].

5. EXPERIMENT

In the first experiment, we tested whether the proposed tool helps reducing users' workload (i.e. cognitive burdens) in exploratory search, the primary motivation for devising the proposed method. In the second experiment, we tested the tool for its usefulness in reusing previously encountered information. More specifically, it tested how the proposed tool helps users in performing tasks that require organizing and remembering the results from searching and browsing.

Experiment 1: Reducing Workload

Our first interest was to find out whether the system implemented based on the two-level model would help reducing workload of users. Given the motivations of our work, workload is a reasonable measurement to test the tool's efficacy because it measures how much effort is required to complete an exploratory search task. In this experiment, we used a special instrument, subjective workload assessment technique (SWAT) [12]. This method has been utilized for evaluating three criteria: time, mental effort, and stress.

We asked the participants to perform a total of 10 exploratory search tasks in the Wikipedia environment where the articles were judged for usefulness in learning background and detailed information for exploratory search tasks. In this experiment, we utilized a simple English Wikipedia, and evaluated efficacy of our information seeking interface as an aid to exploratory search. Each task has one topic selected from the topics of 10 different Wikipedia categories. For a more realistic exploratory search environment, we provided blank forms that they had to fill out. The forms are composed of two parts: semantic annotation and summarizing. Semantic annotation is to annotate information about what related entities appear in texts, and summarization means answering non-factoid questions like "writing a state of the art" and "writing important background information". To minimize potential biases like leaning effects, the participants applied two methods, with and without the interface, in an alternating fashion.

Table 1. The result of SWAT

	with interface (Average SD)	without interface (Average SD)	Difference
Time	1.6 (0.55)	1.8 (0.45)	+ 0.2
Mental effort	1.2 (0.45)	2.4 (0.55)	- 1.2
Stress	1.8 (0.45)	2.2 (0.45)	- 0.4
Total	4.6 (0.89)	6.4 (0.89)	- 1.8

The participants' rates of SWAT range between 1 (the best) and 3, and the result of workload analysis is presented in Table 1. Our interface received a mean score of 4.6, which is a significant improvement over the case without the interface. In particular, the difference was the greatest for mental efforts as intended and expected for the interface. These observations showed that our new information seeking interface helped reducing workload in three different ways in the task of exploratory search.

Experiment 2: Information Reuse

Since our two-level model and its manifestation as a tool were devised to help users reducing cognitive efforts in information seeking processes, manifested by searching and browsing activities, we decided to focus on information reuse activities in information seeking. In the web environment, users often have to skim through an overwhelming amount of information, suffering from information overload, before their goals are achieved. In this experiment, The three methods, the Favorites tool, SIS [13], and *sketchBrain*, were compared in six different tasks by ten groups of users, each consisting of three undergraduate students. In total, 30 users were employed for six different tasks using three different methods. The six tasks consist of questions in six different domains like Medicine and Sports. The tasks were designed as follows. For a task, the participants (users) were first asked to read 30 pre-selected web pages. One minute per page was given to simulate an information skimming situation. The participants were then asked to organize the pages using the given tool within one minute. After the preparation stage, they were given three information hunting questions elicited from the 30 pages they read. The participants were timed for completion of each question answering. Since the maximum time given to each question was five minutes, the time taken for an unsolved question was assumed to be solved in five minutes, the maximum. In order to minimize user dependency and learning effects, the users were assigned to six tasks using three different methods in an alternating fashion. Each user evaluated each method twice for different tasks, and each task was given to the three users in an effort to minimize user dependency. Three users used the three methods in different sequences for different tasks so that there is little learning effect on average.

To ensure that every participant has some familiarity with the three tools, we gave them a tutorial with 10 minutes of practice sessions in the same place with all the participants together.

Table 2. ANOVA result

			95% Confidence Interval for Mean	
Methods	Mean	Std.Deviation	Lower Bound	Upper Bound
1: Favorites	87.69	98.82	62.16	113.22
2: SIS	70.09	67.67	52.61	87.57
3: Our Tool	50.33	43.78	39.02	61.64

The comparison result is shown in Table. 2. It took about 50 seconds on average to solve the problems using our tool, but 88 (about 76% longer) and 70 seconds (about 40% longer) using the Favorites tool and SIS, respectively. Although SIS didn't require any extra user efforts to organize the pages, the time spent on the organization was only one minute, once for all the tasks. If the initial investment for our tool is spread across all the questions, the extra time spent is very small. In ANOVA analysis, it shows that the mean for our tool was better than those of Favorite and SIS. ANOVA puts all the data into one number (F) and gives us one P for the null hypothesis. The value was equal to $F(2,177)=3.866$ ($p < 0.05$), and the difference was reliable at the 95% confidence level. It means that users were more likely to say that our tool had superior information reusability.

6. CONCLUSION

We have proposed a new information seeking interface for extracting/utilizing cognitive personal knowledge structure, which explicates operations at the knowledge level and across the information and knowledge spaces in addition to the typical information level operations, searching and browsing. The tool we developed, which is a limited manifestation of the model, was first tested how the tool is helpful to reduce cognitive burden. Based on the encouraging results, we conducted a more focused and carefully designed experiment to evaluate the tool's utility in reusing a relatively large amount of information that has been encountered. In comparison, our tool was superior to the others in supporting information reuse tasks. The result indicates that our novel approach, the two level model and the associated operations, is very promising and worth further study. First of all, the two-level model can be extended further and implemented in other ways with different emphases. For example, it would be useful to search using a topic-association-topic triplet as a query. In this case, information objects need to be indexed accordingly. Second, automatic generation of topics and associations require further research, which is essential to reducing users' burden in constructing their own knowledge space. Finally, a complete system based on the two-level model must be deployed to a real user environment for more extensive experiments.

7. REFERENCES

[1] Bates, M. 1989. The design of browsing and berry picking techniques for the online search interface. Online Re-view. 13, 5, 407-431.

[2] Pirolli, P. & Card, S.K. 1995. Information foraging. *Psychological Review*, 106, 643-6753.

[3] Russell, D.M. et al. 1993. The cost structure of sensemaking, Proc. of CHI 19*93.* 269-276.

[4] Marchionini. G. 2006. Exploratory search: from finding to understanding. Commun. ACM. 49, 4, 41-46.

[5] White, R. W. and Drucker, S. M. 2007: Investigating behavioral variability in web search. Proc. of WWW 2007. 21-30.

[6] Hendry, D. G., and Harper, D. J. 1997. An informal information-seeking environment. Journal of the American Society for Information Science. 48, 11, 1036-1048.

[7] Buchanan, G., et al. 2002. Spatial Hypertext as a Reader Tool in Digital Libraries. LNCS 2539, 13-24.

[8] Martin, I., and Jose, J.M. 2004. Fetch: A personalized information retrieval tool. Proc. of RIAO'2004 Con.

[9] Harper, D. J. and Kelly, D. 2006. Contextual relevance feed-back. Proc. of IIiX 2006, 129-137.

[10] Park, H., et al. 2007. Personalized Knowledge Structure for Exploratory Search and Information Reuse, http://ir.icu.ac.kr/~phg/snb_techReport1.pdf.

[11] Park, H., et al. 2008. Interactive Information Seeking Interface for Exploratory Search. Proc. of ICEIS 2008. 276-285

[12] Reid, G.B. and Nygren., T.E. 1988. The subjective workload assessment technique: A scaling procedure for measuring mental workload. P.A. HANCOCK and N. MESHKATI, eds. Human Mental Workload. Amsterdam: North Holland.

[13] Dumais, S., et al. 2003. Stuff I've seen: a system for personal information retrieval and re-use. Proc. of SIGIR 2003, 72-79.

[14] Taylor, R. 1968. Question-Negotiation and Information Seeking in Libraries. College & Research Libraries, 29, 3, 178-194.

[15] Teevan, J., et al. 2004. The perfect search engine is not enough: a study of orienteering behavior in directed search. Proc. of CHI 2004, 415-422.

[16] McKenzie, B. and Cockburn, A. 2001. An empirical analysis of web page revisitation. Proc. of HICSS 2001, CD Rom.

[17] Belkin, M.J. 1980. Anomalous states of knowledge as a basis for information retrieval, Canadian Journal of Information Science. 5, 133-143.

[18] Shen, X., Tan, B., and Zhai, C. 2005. Context-sensitive in-formation retrieval using implicit feedback. Proc. of SIGIR 2005, 43-50.

[19] Zigoris, P. and Zhang, Y. 2006. Bayesian adaptive user profiling with explicit & implicit feedback. Proc. of CIKM 2006, 397-404.