Research Paper

A preliminary study on exploratory search behavior of undergraduate students in China*

Yunqiu ZHANG†, Wenxiu AN & Jia FENG

School of Public Health, Jilin University, Changchun 130021, China

Received Jan. 5, 2012
Revised Mar. 28, 2012
Accepted Apr. 26, 2012

Abstract

Purpose: This study attempts to investigate how a user's search behavior changes in the exploratory search process in order to understand the characteristics of the user's search behavior and build a behavioral model.

Design/methodology/approach: Forty-two matriculated full-time senior college students with a female-to-male ratio of 1 to 1 who majored in medical science in Jilin University participated in our experiment. The task of the experiment was to search for information about "the influence of environmental pollution on daily life" in order to write a report about this topic. The research methods include concept map, query log analysis and questionnaire survey.

Findings: The results indicate that exploratory search can significantly change the knowledge structure of searchers. As searchers were moving through different stages of the exploratory search process, they experienced cognitive changes, and their search behaviors were characterized by quick browsing, careful browsing and focused searching.

Research limitations: The study used only one search topic, and there is no comparision or control group. Although we took search habits, personal thinking habits, personality characteristics and professional background into account, a more detailed study to analyze the effects of these factors on exploratory search behavior is needed in our further research.

Practical implications: This study can serve as a reference for other researchers engaged in the same effort to construct the supporting system of exploratory search.

Originality/value: Three methods are used to investigate the behavior characteristics during exploratory search.

Keywords Exploratory search; Search behavior; Concept map; Log analysis

CJLIS
Vol. 5 No. 1, 2012
pp 70–84
National Science Library, Chinese Academy of Sciences

1 Introduction

With the development of computers, the Internet and communications technology, searching on the Internet has become a lifestyle. The classic question and answer

* This work is supported by the National Social Science Foundation (Grant No.: 11BTQ045).
† Corresponding author: Yunqiu Zhang (E-mail: yunqiu@jlu.edu.cn).

information retrieval (IR) model is the dominant interaction model currently used in major commercial Web search engines. However, this model describes the information search process from the information system's perspective instead of the user's perspective. It cannot truly reflect the behavioral changes in the interaction process between the information system and the user or express the changing information need in the information search process (ISP). What is more, it ignores the effects of some significant factors, such as the user's background knowledge about the search topic and the usage of the information. Thus, there is a pressing need for the study of complex information needs and the information search behavior related to the Internet, as well as for the construction of a new type of Web search engine.

To some degree, any actual information needs and search behaviors are complex and exploratory. In the past three decades, many researchers have studied information search behavior, with some focusing on exploratory search behavior. In 1989, Bates proposed a berrypicking model in her paper[1]. She likened finding information to picking huckleberries or blueberries in the forest, from where they are scattered on bushes, rather than clustered in bunches. People must pick the berries one by one. Information searchers move through an information space to gather information. New information may yield new ideas so that searchers will develop a new conception of the query. They may then change their search direction, and as the search progresses, the desired outcome may also change. Bates described this approach as an "evolving search". She stated that the berrypicking model tries to reveal the role of browsing in the search process, because users must scan the results to select the information which satisfies them. In 2007, Bates demonstrated the relationship between berrypicking and browsing once again. She stated that clicking and scanning the search results can be understood as part of the information retrieval strategy. However, berrypicking emphasizes an evolving query, which shifts during the course of the search, rather than the search behavior itself[2]. In 1991, Kuhlthau proposed an ISP model which includes six stages: Initiation, selection, exploration, formulation, collection, and presentation[3]. Yet, the model mainly discusses a user's mood changes in each stage. In 1995 and 1999, Pirolli and Card's study attempted to explain information seeking behaviors in humans by presenting the information foraging theory. Central to the theory is the idea that the mechanism of information foraging is similar to that of food foraging in living organisms. When moving between information patches, searchers may select the paths to maximize the rate of the gain of information relevant to their task[4–5]. The theory is based on the premise that users' information needs are clearly defined in their minds. In 1992

and 1998, Dervin carried out studies and presented the sense-making theory. This theory deems that each individual is an entity, moving through time and space, and the individual can perceive the gaps between entities, time and spaces from time to time in the real world. Only by using information seeking for bridging these gaps can the ideal cognitive results be obtained and the new cognitive state be reached[6–7]. In 1992, Zhang studied university students' information search process, and explored their information seeking behavior and their psychological changes during the search process by giving them three different types of questions with different levels of complexity, analyzing their search log files and interviewing them. The students with different states of knowledge are randomly sampled and then required to perform search tasks. The results were classified into three categories: 1) Students who had specific answers, 2) those who found related information, and 3) those who retrieved some materials, but were uncertain whether or not these materials were useful to answer the questions[8]. In 1995, Byström and Järvelin analyzed the complexity, information type, information path and source of search tasks based on log analysis and questionnaire survey[9].

In 2006, Marchionini proposed "exploratory search"[10]. In 2009, White and Roth stated the characteristics and critical problems of exploratory search, which they believed involved two intellectual activities, namely, learning and investigation. People who are engaged in exploratory searches are generally: 1) Unfamiliar with the domain of their goals; 2) unsure about the ways to achieve their goals; and/or even 3) unsure about their goals[11]. The presentation of the concept improved our understanding of the complicated information search behavior under network environment, and exploratory search gradually became a major focus of study in information retrieval, human-computer interaction, cognitive science, and sociology. Nowadays, the critical problems and challenges in this domain include the construction of the behavioral model and supporting search system and the effective evaluation method and technology. Of these, the behavioral model is the foundation. Based on the background, our study attempts to conduct a preliminary research of the exploratory search behavior through a user experiment combined with other methods.

2 Methodologies

2.1 Research questions and purposes

This study is centered on search behavior which is one of the key issues of exploratory search. According to the analysis of the features of exploratory search, the

National Science Library, Chinese Academy of Sciences

search process involves two kinds of intellectual activities—learning and investigation, which means that a searcher's conception of a topic may change in the course of his or her exploratory search. But how it will change specifically and to what degree it will change and what are characteristics of this change? And how information searchers perform their searches specifically when these changes occur? What are the characteristics of their search behaviors? These are questions that we will delve into in our research. An experiment is designed to analyze the changes of knowledge representations of the searchers by comparing concept maps depicted before and after each search process and analyze the search behavior by using query logs. The study aims to investigate the characteristics of the user's behavioral changes in the exploratory search process and to build a behavioral model.

2.2 Research methods

In our study, a concept map, a tool to assess the learners' knowledge and understanding, is used to assess the changes of information searchers' knowledge structure. To analyze the observable changes in the searchers' behaviors in exploratory searches, we have adopted searchers' query log analysis. In addition, the questionnaire survey is used as an auxiliary method.

2.2.1 Concept map

A concept map is a graphical representation of the concepts of a theme and their relationships, and it is a tool used to organize and represent knowledge. Fig. 1 shows an example of a concept map about wood. The concept map consists of concept words, arrows that connect the concept words, and linking words on the arrows[12]. Concept words are nouns which usually represent events or objects. In the concept map, concept words are generally enclosed in circles or boxes, called nodes. Linking words, also called link labels which are marked above the arrows, are usually constituted by verbs, adjectives, or conjunctions. Arrows, also called links, express the relationships between concept words. Generally speaking, concept words connected by arrows and linking words form sentences, and one example is "wood is used to make paper".

At present, concept maps have been used in the field of educational research, and many studies have shown that concept maps can be used to evaluate learners' knowledge and understanding[13]. However, in the field of information retrieval, there were few published studies focusing on using concept maps to explore the changes in users' knowledge structure during their exploratory search on the Web. In consideration of the complexity of the exploratory search process involving learning and understanding, this study attempts to discuss the changes in the user's knowledge structure on the search task topic during the exploratory search process by analyzing the changes between the pre- and post-search concept maps.

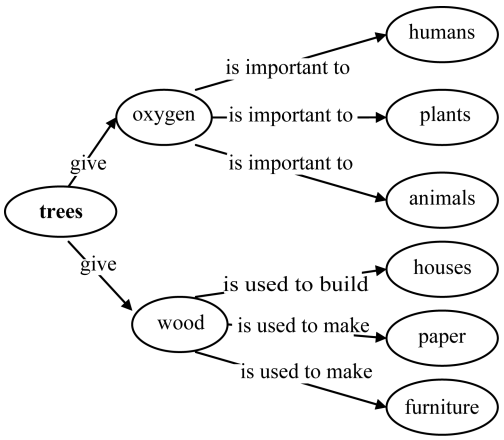

Fig. 1 Example of a concept map about wood.

2.2.2 Log analysis

During the Web search process, users will leave their traces, query log. The query log can objectively record a user's access time, Web pages visited, time spent, questions posed, and query changes. Analysis of the query log can objectively describe an information searcher's exploratory search path and interaction with the system. Therefore, the log analysis has always been the main method used to study users' search behaviors. In this study, with the help of the network and query log files recorded through the Firefox Web browser, we analyzed the search tools, queries and number of pages browsed in the search process. Meanwhile, in order to test how subjects explore the Internet in the search process, the link depth is defined in our experiment. Link depth is defined as an index which indicates the extent to which a searcher has moved deeper into the Web following links from search engine result pages. As shown in Fig. 2, the link depth for a page reached through a link from a search engine result page is 1. The link depth for a page reached through a link from a page with a link depth of 1 is 2. Depth increases by 1 for each link followed.

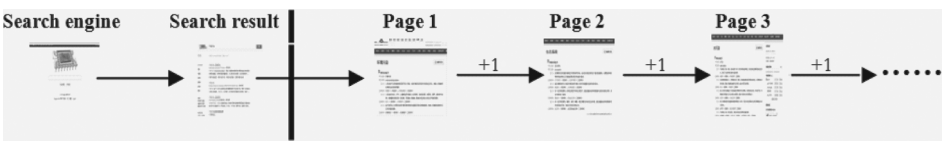

Fig. 2 Example of the link depth.

2.2.3 Questionnaire investigation

Pre- and post-search questionnaires were designed. The pre-search questionnaire was used to collect some general information on the subjects, including their level of understanding of the search topic, the main methods they use to access daily information, their search habits, the commonly used search tools, and the average time they spend online.

The post-search questionnaire included four aspects: 1) Subjects' levels of understanding of the search topic and the changes in their interest; 2) behavioral changes during the search process; 3) factors affecting the search; and 4) degree of satisfaction with the existing search tools in meeting their information needs. The questionnaires included multiple choice questions and open-ended ones. At the end of the questionnaire, subjects were asked to write about the changes in their thoughts during the information search process. The aim of the questionnaires was to reveal the searcher's subjective feelings, thoughts, and behavioral changes during the search process from the searcher's perspective in order to help us understand the results of the concept map and the log analysis.

2.3 Experimental design

2.3.1 Subjects and search task

Because the exploratory search is a high-level search activity, the study chose subjects who have academic backgrounds. To simplify the experiment process and make it operational, the subjects were matriculated full-time senior college students who majored in medical science in Jilin University, China. In all, there were 42 subjects with a female-to-male ratio of 1 to 1.

The task of the experiment was to search for information about "the influence of environmental pollution on daily life" in order to write a report about this topic. The search task was open-ended, which represents the feature that a user can expect from exploratory search.

2.3.2 Procedure

Fig. 3 shows the procedure of the experiment. Firstly, the subjects were given the experiment instruction. They were trained to draw a concept map and asked to practise drawing the map. Next, before searching, they were given 15 minutes to answer the pre-search questionnaire and to draw a pre-search concept map for the assigned topic. After drawing the concept map, they conducted the exploratory search for 30 minutes. Finally, they drew a post-search concept map and finished the post-search questionnaire.

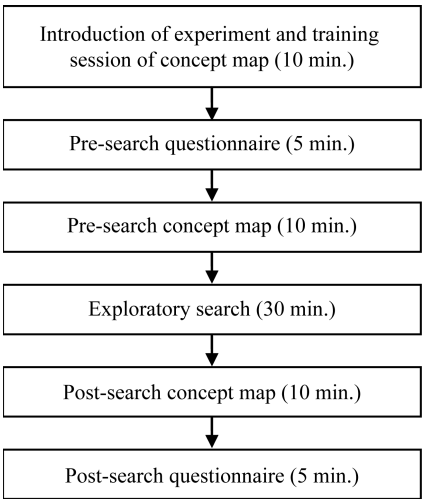

Fig. 3 Experiment procedure.

3 Results

According to the experimental methods of the study, the results were analyzed by using data collected from concept maps, log files and questionnaire survey.

3.1 Concept map

An analysis of a concept map mainly includes 3 groups of indicators: 1) The numbers of nodes, links and link labels, 2) common, lost and new nodes, links and link labels, and 3) the hierarchical structure and features of pre- and post-search concept maps.

3.1.1 Analysis of the number of nodes, links, and link labels

Table 1 lists the minimum, maximum, mean, median and standard deviation of each component of the concept maps. Compared to the pre-concept maps, nodes, links and link labels increase obviously in post-search concept maps. With the *t*-test in components between pre- and post-search concept maps, nodes: t=–9.34, v=40, p=.000; links: t=–9.55, v=40, p=.000; links labels: t=–7.27, v=40, p=.000. It is clear that the change has remarkable statistical significance. In the case of the minimum nodes, there are 4 and 12 nodes in the pre- and post-search concept map, respectively. In terms of the maximum nodes, there are 27 and 65 nodes, respectively. From the perspective of the minimum, maximum, mean and median, changes in the number of nodes and links show a similar tendency. The tendency can be explained by the fact that two nodes usually have a single link and each node is associated

with at least one link. However, the number of link labels has a different distribution from that of the nodes and links. This is because many links are described without labels, with the portion of links having labels at 65.35%. But the general change tendency is consistent.

Table 1 Number of nodes, links and link labels in pre- and post-search concept maps

		Min.	Max.	Mean	Median	SD
Nodes	Pre-node	4	27	13	12	5.18
	Post-node	12	65	28.57	25.5	12.97
Links	Pre-link	4	28	13.71	12.5	6.03
	Post-link	12	60	28.57	26.5	12.51
Labels	Pre-label	0	27	7.57	6	7.2
	Post-label	0	60	17.74	14	14.43

Judging from the numbers of nodes, links and link labels, there are salient changes between the post-search concept maps and the pre-search concept maps, which indicate that the knowledge structure of the subjects changed significantly after the search process. However, judging from the values of standard deviation, individual factors such as search habits and personal thinking habits exerted an increasing effect on the changes of post-search concept maps.

3.1.2 Analysis of common, lost and new nodes between pre- and post-search concept maps

To explore the features of the post-search concept maps, we analyzed common, lost and new nodes between pre- and post-search concept maps. From the comparison, as shown by Table 2 below, the minimum number of common nodes is 0, the maximum number is 15, and the average is 3.38, taking up 26% of average nodes of pre-search concept maps and 11.83% of average nodes of post-search concept maps, respectively. When the post-search concept maps were formed, it is found that the average number of lost nodes was 9.62, accounting for 74% of average nodes of the pre-search concept maps. Compared to the pre-search concept maps, the average number of new nodes of the post-search concept maps was 25.19, accounting for 88.17% of average nodes of the post-search concept maps. In the pre- and post-search concept maps, the minimum number of links is 0, the maximum is 13, and the average number is 2.21, accounting for 16.12% of average links of the pre-search concept maps, and 7.74% of average links of the post-search concept maps, respectively. Until the post-search concept maps were formed, the average number of lost links of the pre-search concept maps was 11.5, accounting

for 83.82% of the pre-search concept maps. Compared to the pre-search concept maps, the average number of post-search new links was 26.36, accounting for 92.26% of average links of the post-search concept maps. In the pre- and post-search concept maps, the minimum number of common link labels is 0, and the maximum is 13, and the average is 1.52, accounting for 20.08% and 8.57% of average pre-search link labels and post-search link labels, respectively. When the post-search concept maps were formed, the average number of lost link labels was 6.05, accounting for 79.92% of average link lables of the pre-search concept maps and the average of new link labels in post-search concept maps was 16.21, accounting for 91.43% of average link labels of the post-search concept maps.

Table 2 Common, lost and new components between pre- and post-search concept maps

		Min.	Max.	Mean	Median	SD
Nodes	Common nodes	0	15	3.38	2	3.27
	Lost nodes	3	19	9.62	8.5	4.58
	New nodes	5	60	25.19	23	12.37
Links	Common links	0	13	2.21	0.5	3.49
	Lost links	4	21	11.5	12	5.05
	New links	8	57	26.36	25.5	11.52
Labels	Common labels	0	13	1.52	0	2.98
	Lost labels	0	19	6.05	5	5.59
	New labels	0	48	16.21	14	12.85

In summary, there were few common nodes and relatively many lost and new nodes. This indicates that subjects were slightly influenced by their previous knowledge, while they were endeavoring to form a new knowledge structure.

3.1.3 Analysis of hierarchical structure and features of pre- and post-search concept maps

The previous analysis was merely focused on the basic components of pre- and post-search concept maps. Considering that subjects experienced some degree of cognitive changes during the search process, we also studied the hierarchical structure of concept maps. The number of the tier of the hierarchical structure of the post-search concept maps has increased by 137 and the average was 3.26. These changes to the hierarchical structure were not remarkable. Nevertheless, the obvious increase in the number of nodes indicates that subjects tried to specify the related knowledge in the same hierarchical structure during the search process. This

indicates characteristics of divergent thinking. A total of 40 subjects started their search with the phrase "environmental pollution" and formed the first 2 to 3-tiered hierarchical structure of their concept maps. The post-search concept maps of 29 subjects (70% of the total number) showed the refining in particular branches of the structure. As Fig. 4 shows, there is a concentrated thinking which reveals that subjects had a focused interest. In addition, all concept maps consisted of nodes related to diseases, which may be due to the academic background of these subjects rather than the search task.

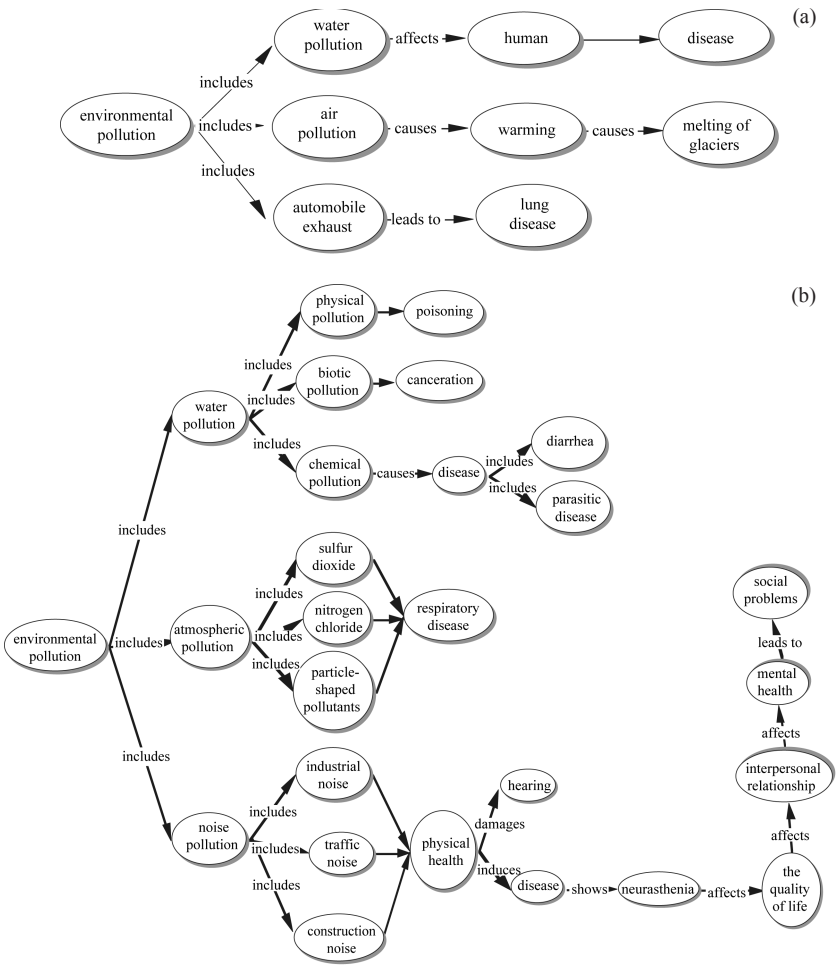

Fig. 4 (a) Pre-search concept map; (b) Post-search concept map.

Research Paper

The pre-search concept maps are simple, lacking systematic structure or accuracy, whereas the post-search concept maps are obviously complicated, systematic, professional and accurate.

3.2 Log analysis

Our study analyzed the search tools, queries, number of pages and depth of links of subjects during their search process by using the browser log files through the Firefox Web browser.

3.2.1 Search tools and queries

According to the log files, the main search tools used by subjects were Baidu, Google, CNKI journal full-text database, VIP journal full-text database, Sina, NetEase and Docin.com. This corresponded with the search habits of these subjects, who preferred some commonly used search tools.

All of the subjects used the query term "environmental pollution". As Fig. 5 shows, there were 59 terms related to "environmental pollution", including the following 4 categories: 1) Terms directly related to the task, including daily life, influence, harm, prevention, human and human health, etc., 2) terms related to specific types of environmental pollution, including noise pollution, water pollution, light pollution, marine pollution, radioactive contamination, chemical pollution, industrial pollution and agricultural pollution, etc., 3) terms related to actual harm caused by environmental pollution, including cancer, deformity, poisoning, neurasthenia, emphysema, hearing loss, etc., and 4) terms related to social problems caused by environmental pollution, including the quality of life, school life, social problems, society responsibility, social cost, interpersonal relationship, etc.

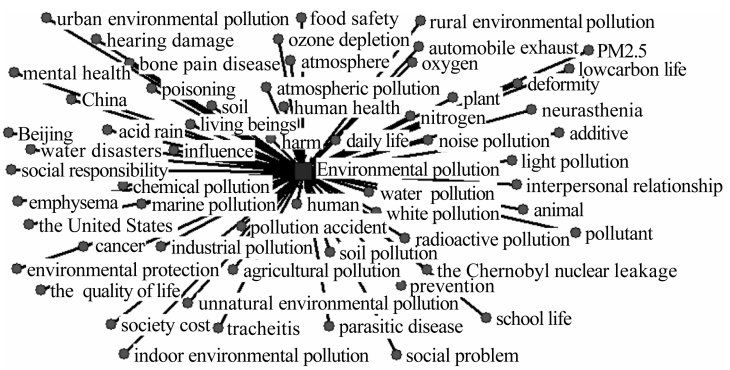

Fig. 5 Query terms in search process.

From the results, the first 3 categories of search terms reflected similar ways of thinking. Based on the preliminary understanding of the task, the subjects started their search from "environmental pollution", explored various specific types of environmental pollution, and found the impact of environmental pollution on people's daily lives. The fourth category of search terms reflected the personal interest of subjects in the search process.

3.2.2 Pages and link depth

According to log files, we performed an analysis on the numbers and the link depth of Web pages. The results showed that the number of Web pages browsed during the search process ranged from 5 to 86, with an average of 32.05. Because the time given to subjects to carry out exploratory search tasks was 30 minutes, the average time spent browsing a website was 56.16 seconds. To analyze the exploration paths of each subject, the depth of links is 1 to 3 in the process of searching.

Fig. 6 is an example of the link depth of one subject. It shows that the search process contains three kinds of behaviors: 1) Quick browsing. It refers to browsing the results of broad query terms, which shows that the subject explored one page downward from the search results, and then went back and clicked another link. The link depth is 1. 2) Careful browsing. It refers to browsing the results returned after search engine searched one specific query term. The depth is 1 to 3. 3) Focused searching. It means that the searcher browsed the results after search engine searched

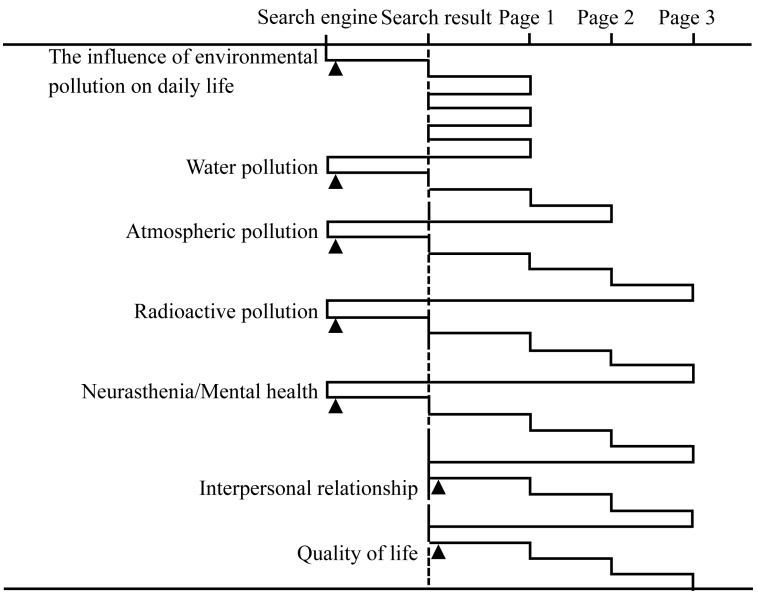

Fig. 6 Example of link depth (▲ represents entering the query words).

several specific terms. During the sub-process of searching, the subject added new specific terms to the query. The link depth is 1 to 3.

3.3 Analysis of questionnaire results

According to the pre-search questionnaires, all the subjects used the Internet as a means of obtaining information. The most commonly used search engines were Baidu and Google. With about 30 minutes of Web searching a day, the subjects have some awareness of information literacy and some search skills. They had some knowledge about the search topic, but did not have specialized knowledge.

The results of the post-search questionnaires revealed that all the subjects believed that their knowledge of the search topic increased after the search. 21.43% of subjects agreed that the improvement was significant, 78.57% of subjects stated that their knowledge improved to a certain degree. Finally, 78.57% of the subjects stated that their interest in the search topic increased. All the subjects conducted extended retrieval when they discovered related knowledge during the search process. 16.67% of them extended the search until they completely understood the search topic. 83.37% of them extended the search only to a certain degree, and then they changed direction once they got a rough idea of the search topic. According to these subjects' opinions, the main factors which influenced the process of exploratory search were ranked as follows: Specialized background knowledge, thinking habits and personality characteristics, and search habits. 88.1% of the subjects held the view that the existing search tools needed to be improved in order to better support exploratory search.

4 Conclusions

By comprehensively analyzing concept maps drawn before and after the exploratory search, the query logs and questionnaire results, we can describe the behavioral patterns of exploratory search (Fig. 7). Along with cognitive changes during the search process, information searchers' behaviors change from quick browsing to careful browsing and finally to focused searching.

As searchers are moving through different stages of the exploratory search process, they exhibit different search behaviors and experience cognitive changes. When faced with a task to search an unclear topic in an unfamiliar field, due to a lack of knowledge on the topic, searchers firstly seek a preliminary understanding of the information through quick browsing. From their own personal search habits, searchers choose their own search tools and broad search words. The search link depth is 1, and the searchers return to browse the results for several times. At this point, the searchers' knowledge has increased in the short-term memory. On this basis, searchers choose the easiest sub-subject to form the knowledge system.

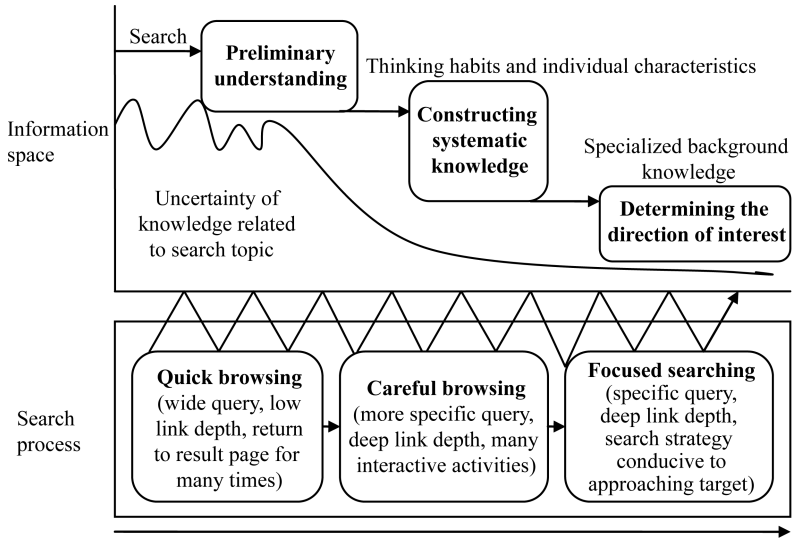

Fig. 7 Behavioral model of exploratory search.

Trying to build a new knowledge structure, they take in, analyze and synthesize each piece of information. This represents careful browsing. Information searchers continue to repeat the process with more specific search terms and link depths between 1 and 3, which is affected by their thinking habits and personalities. As the exploration process continues, they grow less uncertain of the search topic, and their understanding has improved. That is to say, information searchers are gradually able to understand the overall situation and become aware of their positions in space. Meanwhile, they become clear of their information needs, and they can judge whether the information is suitable or not. Influenced by their previous background knowledge, their specific interests begin to come together, and the search goal is clear, and they start focused searching. At this stage, the information searchers adopt traditional search strategies, using more specific search terms, and gradually trying specialized terms. During the process, they are approaching their target.

In summary, we conducted a research on exploratory search behaviors based on the fact that the subjects were college students with certain search capabilities and we assumed that their aim was to write a report related to the topic. We also described and analyzed the common cognitive changes and the changes to the search terms. Although we took search habits, personal thinking habits, personality characteristics, and specialized background knowledge into account, we have not analyzed the effects of these factors in details. Therefore, the effects of different search tasks, user types, as well as subjective factors on search behaviors in the exploratory search process will be investigated in our future research.

Research Paper

References

1 Bates, M. J. The design of browsing and berrypicking: Techniques for the online search interface. Online Review, 1989, 13(5): 407–424.

2 Bates, M. J. What is browsing – really? A model drawing from behavioural science research. Information Research, 2007, 12(4): paper 330.

3 Kuhlthau, C. C. Inside the search process: Information seeking from the user's perspective. Journal of the American Society for Information Science, 1991, 42(5): 361–371.

4 Pirolli, P., & Card, S. Information foraging in information access environments. In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems. New York: ACM Press, 1995: 51–58.

5 Pirolli, P., & Card, S. Information foraging. Psychology Review, 1999, 106(4): 643–675.

6 Dervin, B. From the mind's eye of the user: The sense-making qualitative–quantitative methodology. Qualitative Research in Information Management, 1992: 64–81.

7 Dervin, B. Sense-making theory and practice: An overview of user interests in knowledge seeking and use. Journal of Knowledge Management,1998, 2(2): 36–46.

8 Zhang, X. L. Information-seeking patterns and behaviors of selected undergraduate students in a Chinese university. Ann Arbor, Mich.: University Microfilms International, 1992.

9 Byström, K., & Järvelin, K. Task complexity affects information seeking and use. Information Processing and Management, 1995, 31(2): 191–213.

10 Marchionini, G. Exploratory search: From finding to understanding. Communications of the ACM, 2006, 49(4): 41–46.

11 White, R. W., & Roth, R. A. Exploratory search: Beyond the query-response paradigm. San Rafael, CA: Morgan & Claypool, 2009.

12 Markham, K. M., Mintzes, J. J., & Jones, M. G. The concept map as a research and evaluation tool: Further evidence of validity. Journal of Research in Science Teaching, 1994, 31: 91–101.

13 Edmondson, K. M. Assessing science understanding through concept maps. In Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (Eds.), Assessing science understanding: A human constructivist view (Educational Psychology). San Diego, CA: Academic Press, 2000.