The effects of conceptual description and search practice on users’ mental models and information seeking in a case-based reasoning retrieval system

Wu He a,*, Sanda Erdelez b, Feng-Kwei Wang c, Chi-Ren Shyu d

a Center for Learning Technologies, Old Dominion University, Norfolk, VA 23529, United States
b School of Information Science and Learning Technologies, University of Missouri-Columbia, Columbia, MO 65211, United States
c ITRI College, Industrial Technology Research Institute, Taiwan
d Department of Computer Science, University of Missouri-Columbia, Columbia, MO 65211, United States

Received 31 October 2006; received in revised form 19 February 2007; accepted 3 March 2007
Available online 27 April 2007

Abstract

This paper reports a study that investigated the effects of conceptual description and search practice on users’ mental models and information seeking in a case-based reasoning retrieval (CBR) system with a best match search mechanism. This study also examined how the presence of a mental model affects the users’ search performance and satisfaction in this system. The results of this study revealed that the conceptual description and search practice treatments do not have significantly different effects on the types of user’s mental models, search correctness, and search satisfaction. However, the search practice group spent significantly less time than the conceptual description group in finding the results. Qualitative analysis for the subjects’ post mental models revealed that subjects in the conceptual description group seem to have more complete mental models of the best match system than those in the search practice group. This study also found that subjects with the best match mental models have significantly higher search correctness and search result satisfaction than subjects without the best match mental models. However, the best match mental models do not guarantee less search time in finding the results. This study did not find a significant correlation among search time, search correctness and search satisfaction. The study concludes with suggestions for future research and implications for system developers who are interested in CBR retrieval systems.

Keywords: Mental models; Information seeking; Conceptual description; Search practice; Case-based reasoning

1. Introduction

Traditional searching for information relies on “exact match” search mechanism that encourages the use of Boolean queries and keywords. However, in recent years some retrieval systems have started to incorporate
search engines based on a "best match" search mechanism (Belkin & Croft, 1992). This mechanism relies on unstructured user queries and ranks search results according to their probability of relevance to the query.

A best match search engine often expands or transforms a user’s query during the retrieval process. This transformation process may have a profound influence on both the search results and the users’ ability to understand the relationship between the query and the returned results (Muramatsu & Pratt, 2001). For instance, users may find unexpected search results and get confused. Thus, for users accustomed to traditional keyword match with Boolean operators using a system with a best match search mechanism can be challenging.

The case-based reasoning (CBR) system is a type of best match system and is relatively new for most end users. A large number of successful commercial and industrial applications have been developed and are now in daily use in areas such as customer support, sales support, diagnostics and help-desk systems (Watson, 1997, 2003). The basic idea of CBR is to suggest new solutions by adapting old solutions to new problems (Aamodt & Plaza, 1994). When faced with a problem, people typically rely on their concrete past experiences in solving problems. For example, given a target problem, many people will assess various situational features presented in the problem, retrieve past experiences from memory with similar situations, and then apply the lessons from these experiences to develop new solutions. Thus, the idea of creating CBR systems is to imitate human expert’s reasoning and use a database of problems with solutions as an external memory. When encountering a new problem, people can search the CBR system, retrieve a problem situation similar to the current one and then adapt its solutions for use in solving the current problem. According to Kolonder (1993), CBR has proved to be an extremely effective approach in solving complex problems. However, for many new users, learning about the use of CBR systems could be a challenging task.

An example of CBR systems is the Knowledge Innovation for Technology in Education (KITE) system, the first developed CBR system in the educational technology integration community. The KITE project at the University of Missouri-Columbia seeks to assist teachers by providing access to a case library with over 1000 stories of teachers’ experiences with technology. The KITE system addresses the challenge of how a case library can be used to support pre-service and in-service teachers who are learning how to integrate technologies into their teaching (Wang, Means, & Wedman, 2003a). Built on the idea of case-based reasoning, the KITE system implemented a CBR search engine. The goal of the KITE CBR search engine is to retrieve cases based on semantic meanings (similarities) of cases. As illustrated in Fig. 1, there are four major components in

Fig. 1. KITE CBR Search Engine’s Architecture (adopted from Wang et al., 2003b).
the KITE CBR retrieval system: (1) the case library, (2) the feature vector space, (3) the user interface, and (4) the search engine.

Technically, the KITE search engine will automatically expand the user’s search terms by adding all the other terms within the same field. These added terms have different weight or importance, depending on their similarity to the concept of the user’s search terms. When the search is complete, the search engine will first return cases that match all search criteria selected, followed by cases that closely match the users’ search criteria. If there are no exact matches, the search engine will provide the user with a list of ranked cases that best match the user’s search criteria. For example, as Jonassen and Erdelez (2005) described below:

...a search for stories related to K–1 grade level may return cases related to grade 2–3 and a search for stories about teachers’ use of digital camera may also return cases that report on the use of video cameras. In the above examples, the CBR system will automatically recognize that grades 2–3 are close to grades K–1, and that a use of a digital camera is conceptually close to a use of a video camera (P. 5).

2. Statement of the problem

Best match systems such as CBR retrieval systems are increasingly common in academic environments. However, finding information in best match systems is becoming more challenging as a greater numbers of people with diverse user styles and backgrounds are able to access these systems.

For example, Wang, Moore, Wedman, and Shyu (2003b) found that “navigation of the KITE CBR retrieval system by users without prior CBR experience is often a daunting task.” As the KITE CBR retrieval system often expands or transforms a user’s query by adding terms similar to the concept of the user’s query during the retrieval process, many novice users get confused with the search results and are unable to determine the relationship between the query and the returned results. Muramatsu and Pratt (2001) also found in a search engine study that many users were frustrated because they did not understand how the search engines were transforming their queries.

According to researchers such as Wang et al. (2003b), Muramatsu and Pratt (2001) and Jonassen and Erdelez (2005), part of the above problem may be caused by the lack of experience and appropriate mental models. The evaluation study on the KITE CBR system conducted by Wang et al. (2003b) showed that most participants’ information search experiences were strongly rooted in conventional information retrieval systems that rely on a Boolean or keyword search mechanism. These novice users have difficulties in discerning the differences between a CBR search and a keyword search. Furthermore, researchers (Dimitroff & Wolfram, 1995; Jonassen & Erdelez, 2005) suggested that users’ prior experience with traditional Boolean-based information retrieval systems may create a mental barrier in their adoption of new mental models of online searching and, therefore, create frustrations in attempts to use the new system. These past experiences may explain why users feel uncomfortable with the use of the KITE CBR search engine, particularly novice users who had a difficult time adjusting to the new search environment.

To alleviate the problem and foster the use of the KITE CBR search engine, there is a need to better educate the users about how the KITE CBR search engine works. Empirical research is needed to evaluate the effects of training on the users’ use and mental models with the KITE CBR search engine.

3. Literature review

3.1. Fundamentals of mental models

The term mental model generally refers to “psychological representations that aid in understanding, explaining, or predicting how a system works” (Johnson-Laird, 1983; Norman, 1988). Mental models are generally descriptive in nature, describing physical objects and how they work, and are developed through experience, training, and instruction.

The study of mental models poses some methodological issues. As mental models are the internal conceptual representations that people develop while interacting with the environment, it is very difficult to observe,
eliciting, and measure them. People normally have difficulty articulating the structure and content of their own knowledge of the systems they interact with (Briggs, 1987; Norman, 1983; Zhang & Chignell, 2001).

Despite the methodological difficulties of capturing a person’s mental model, researchers have explored mental models in many ways. As a user’s mental model is also embedded in the activities and artifacts engaged in by the user (Jonassen, 1999), Choo, Detlor, and Turnbull (2000) attempted to identify characteristics of mental models related to searching. Their results revealed that aspects of mental models can be interpreted in terms of search behavior. Choo et al. (2000) suggest that a detailed analysis of users’ search strategies and searching tactics will help improve the understanding of mental models of information seeking. Some other studies also have shown that a user’s interaction with a given computer system is guided by the user’s mental model of that system (Borgman, 1986; Carroll & Olson, 1988; Preece et al., 1994). In other words, the mental models formed about a system can influence a user’s performance and behavior (Baecker & Buxton, 1987). Another study by Savage, Belkin, Cool, and Xie (1997) suggests that mental models allow people to infer how the system works, what kind of input should be provided and what the output means. As mental models can guide users’ operation, expectations, and understanding of the systems, it is useful to examine users’ expectations or views towards a system when investigating the user’s mental model of that system (Zhang & Chignell, 2001). Young (1983) proposes several practical suggestions for researchers to find evidence of users’ mental models. He suggests that researchers should observe how users learn and use the system, how users explain the system and how users predict the behavior of the system.

3.2. Mental model studies in information retrieval systems

There have been a number of mental model studies in the field of information retrieval (IR). Most of these studies investigate how mental models influence users’ performance and searching behavior in interactive information retrieval systems (Borgman, 1986; Choo et al., 2000; Cool, Park, Belkin, & Koenemann, 1996; Dimitroff, 1992; Marchionini, 1989, 1995; Savage et al., 1997; Savage, 2001; Slone, 2002).

For example, Borgman (1986) compared user performance and training methods on an online public access catalog (OPAC), assuming that different training methods would generate different mental models. Her study demonstrated that “people’s mental models of IR can significantly affect their performance and behavior in interaction with an IR system on complex problem-solving tasks”. Dimitroff (1992) studied the relationships between mental models and search performance/outcome. The results of the study show that subjects with more complete models made significantly fewer errors and found significantly more relevant items. Cool et al. (1996) found that mental models may improve search performance. For example, when looking at the number of hits in search results, searchers with appropriate mental models felt more confident of their searching because they knew how to refine the query, based on the current situation.

On the other hand, Savage et al. (1997) explored the relationship between mental models and information seeking in an information retrieval system. The participants’ mental models were classified in relationship to search strategies. Their study reported that searchers possessed various degrees of mental models during the information seeking process. However, this study did not find out how searchers change mental models over time and across various tasks or systems. Furthermore, Wang, Hawk, and Tenopir (2000) conducted analyses of how graduate students in information science programs locate content online. Their study found that many experienced searchers exhibited a number of misconceptions about how a search engine operates. However, they did not find a significant relationship between search performance (measured by search time on task) and the length of system exposure. That is, repeated use of the system does not necessarily improve the search performance if users possess wrong mental models on the system.

In summary, the above studies indicated that “correct” or “faulty” mental models do not guarantee success or failure in search performance.

3.3. Training effects

There are a few training treatments for developing mental models. Two common training treatments are conceptual description and organized search practice. Conceptual descriptions are often used to increase users’ understanding and performance with the use of computer systems (Bayman & Mayer, 1988; Colaric, 2001;
Mayer & Gallini, 1990). Savage (2001) conducted a study to investigate mental model construction and evolution in an IR system. The study found that an in-depth explanation of the internal system operation mechanism led to the construction of a mental model that was more congruent with the system’s operation. Considering the complexity of the CBR retrieval mechanism, a description explaining how the CBR retrieval system works appears to be an appropriate way to help novice users conduct effective searches in a CBR search environment.

On the other hand, several researchers in the field of information science have suggested that system knowledge is best acquired during the use of the system, and that prolonged use increases proficiency (Campagnoni & Ehrlich, 1989; Hsieh-Yee, 1993; Jacobson & Fusani, 1992; Katzeff, 1990). Researchers (Von Glasersfeld, 1995; Jonassen, 1999) also suggests that knowledge and meaningful learning are best built up by the doing and experiencing approach. In particular, Katzeff (1990) found that practice in retrieving information from a database system resulted in increased proficiency and comfort levels in using the system. The study suggests that longer use can increase the users’ understanding of the system. Mayer (1997) reported that users who had a significant amount of experience with the target system through repeated use usually understood the system better than the novices. The novices made many errors initially and needed a lot of time to solve problems when they used an information system. After a period of use, users were able to complete the task much quicker and solve problems in a more efficient way.

4. Purpose of the study

The literature review contributed to our understanding of the effective use of information retrieval systems and provided a framework for the research on CBR retrieval. The authors found few user studies about CBR systems such as user training, users’ mental models, and interaction with the systems. An early work done by Belkin, Cool, Stein, and Thiel (1995) looked at user interaction with a prototype IR system named MERIT. Belkin et al. used a case-based reasoning approach to help structure and organize interaction sequences involving combinations of information seeking strategies in this system. To date, however, no research has investigated the effects of conceptual description and search practice on the users’ construction of mental models using the CBR retrieval systems. Further research is needed to provide evidence on whether a mental model of a CBR retrieval system is best constructed through search practice with the system (system exposure) or through explicit description that presents the CBR retrieval system’s conceptual model to the user.

Moreover, researchers (Borgman, 1986, 1999; Choo et al., 2000; Cool et al., 1996; Dimitroff, 1992; Marchionini, 1989, 1995; Savage et al., 1997; Slone, 2002) have looked at users’ mental models in traditional Boolean information retrieval systems and search engines. For example, they investigated how users’ mental models in Boolean information retrieval systems influence users’ search performance and behaviors in these systems. However, as discussed previously, the traditional IR systems differ from CBR systems in terms of their retrieval mechanism. Boolean information retrieval systems are exact match systems in which search results are usually based on the application of the Boolean logic and the frequencies of query terms. In contrast, the retrieval process of CBR systems uses a best match search mechanism, and search results are ranked on the semantic meanings (similarities) of query terms. For most users, the best match concept of case-based reasoning is relatively new. As no public research to date has investigated the relationships among users’ mental models, search performance and satisfaction in the context of case-based reasoning retrieval systems, it is still not clear whether the research findings acquired in traditional IR systems are applicable to the CBR environments.

The purpose of the study was to investigate the effects of conceptual description and search practice on users’ construction of mental models, search performance (measured by search correctness and search time) and satisfaction in their use of a case library with a best match search mechanism. This study also sought to understand how the presence of a mental model affects the users’ search performance and satisfaction in using the case library with the CBR search mechanism. The KITE case library was used as the CBR application where the participants’ mental models and information seeking behavior were explored.

5. Significance of the study

Many systems using best match search mechanisms have been developed and successfully deployed in academic and business environments. It is important to ensure that users understand that best match systems
operate differently from traditional exact match systems and that they can benefit from using best match systems for effective information searching.

The study makes an important contribution to information science research on users’ mental models of CBR retrieval systems. Because these models are internal conceptual representations that people develop, they are difficult to observe, elicit, and measure. This study uses several complementing data collection methods to develop a presentation, as holistic as possible, of users’ mental models of a case library with a best match search mechanism.

In summary, by investigating the effect of conceptual description and search practice on users’ mental models in a case library using CBR search mechanism, more detailed knowledge can be gained on how case library users construct and adapt mental models through interaction with the best match search environments. The knowledge will help system designers and trainers to create user support mechanisms that will assist users in forming appropriate models for successful utilization of a new generation of information retrieval systems.

6. Research hypothesis

This study examined the following research hypotheses in the context of a case-based reasoning retrieval system:

- H1: There is significant difference between conceptual description and search practice treatments in terms of the effects on the change of the types of users’ mental models.
- H2: There is significant difference between conceptual description and search practice treatments in terms of the effects on search correctness.
- H3: There is significant difference between conceptual description and search practice treatments in terms of the effects on search time.
- H4: There is significant difference between conceptual description and search practice treatments in terms of the effects on search satisfaction.
- H5: There are significant correlations among users’ mental models, search satisfaction, search time and search correctness.

7. Methodology

This study collected data using a combination of both qualitative (interview) and quantitative (questionnaire and search logs) approaches. It is believed that a mixed method approach could offer more accuracy and reliability in information seeking studies (Martzoukou, 2004).

7.1. Participants and settings

Forty subjects were recruited for this study from the undergraduate student body in the College of Arts and Letters at a large public university through a number of course instructors, flyers, and advertisements. An online user screening questionnaire was used to select subjects by identifying demographic information on potential participants such as age, gender, academic background, prior experience in using the PC/Internet/search engines and IR systems, and prior knowledge related to the instructional technology and the CBR system in this study. A total of 64 students filled out the user screening questionnaire. The first 40 students who applied and met all the selection criteria were chosen as subjects for the formal study. Specifically, there were 27 females and 13 males. Thirty-six students were in the age group 18–25 years, and four were in the age group 26–33 years.

Subjects were selected for participation according to the following criteria:

- All subjects were undergraduate students from the College of Arts and Letters.
- All subjects reported feeling (at least) comfortable with computer and Web searching and had experience using the Web on a daily basis.
- All subjects reported no prior experience with using case-based reasoning systems.
- All subjects reported no advanced domain knowledge in instructional technology.
The 40 subjects were randomly assigned to the two treatment groups (Conceptual Description and Search Practice). Each treatment group included half of the participants. In particular, the search practice group included 14 females and six males. Eighteen students were in the age group 18–25 years, and two were in the age group 26–33 years. In contrast, the conceptual description group included 13 females and seven males. Eighteen students were in the age group 18–25 years, and two were in the age group 26–33 years. We also performed a chi square test in SPSS to test the relationships between the treatment assignment and the characteristics of the study sample. There were no significant associations found between the sample characteristics and the treatment that the subjects received. That is, the characteristics of the study sample such as gender, age, computer experience, and search experience relative to the treatment levels appear homogenous in this study.

The data collection process was conducted over a one-month period in a dedicated computer lab. All subjects were tested individually. Each of the 40 subjects received $10 as compensation for their time in the research experiment. Each testing took approximately 45–60 minutes to complete. A screen capture software program was installed in the data collection computer to capture subjects’ interaction with the system. During the study, the screen capture program recorded all the screen displays such as the mouse movement and page navigation while users were doing search tasks. The test-bed for the study was the KITE case library with over 1000 stories of describing teachers’ experiences with technology integration in K-12 classrooms (http://kite.missouri.edu) (Wang et al., 2003b).

### 7.2. Procedures

We administered two different treatments to two groups of 20 randomly assigned study participants. The treatment for Group 1 was text-based conceptual description about the CBR search mechanism implemented in the KITE case library. The conceptual description was evaluated for quality by an experienced instructional designer and two system experts. We ensured that all participants from Group 1 carefully read the printed conceptual description by explaining the purpose of the conceptual description and by asking them to answer content-related questions.

The treatment for Group 2 was an organized opportunity for individual practice with the KITE case library (http://kite.missouri.edu/jkite/ad_search.htm). Each participant in this group was first given an opportunity to explore the search engine on his or her own. Then we assigned each subject six practice search tasks, which were monitored for completion. The six practice tasks specified some known characteristics and required the user to find a case from the case library that best matches the known characteristics. Before they started the practice tasks, we asked each subject to think about how the KITE search engine returns articles to match search criteria during practice. As counting on subjects to experience the best match effect through their own searching could be risky and time intensive, we selected practice tasks requiring queries which would possibly highlight the effect of the query expansion of indexed search terms. This approach allowed subjects to discover the effects of the query expansion of indexed search terms in terms of differential results (Muramatsu & Pratt, 2001). It also guaranteed that subjects experienced the effects of the targeted query expansion and maximized the chance that they would notice the difference in query results. An example practice search question is listed as follows:

“A foreign language subject teacher is interested in using “conferencing/BBS” technology in his lesson to engage students in discussion of books they read. Find a case that best fits this scenario and write down its case number”.

Each user session was broken into two phases (Fig. 2). Phase 1 was designed to explore the effects of treatments on subjects’ mental models. Phase 2 was designed to investigate subjects’ information seeking after receiving their respective treatments.

During phase 1 we:

1. provided subjects a general introduction to the test system,
2. showed a search query to each subject and then asked the subjects to describe, based on prior experience with other search engines, how they thought the search engine would process the query terms and to
predict what kinds of results the search engine would return if executing the query. This explanation and prediction method was first proposed by Young (1983) for researchers to find evidence of users’ existing mental models. Young made the suggestion that researchers should observe how users learn and use the system, how users explain the system and how users predict the behavior of the system. Because some subjects might have difficulty verbalizing comments, the researcher also encouraged the subjects to write down comments or draw pictures to list potential search results on the paper questionnaire. This step was also used as a pre-test to eliminate any of the recruited subjects who already had ideas about best match mechanisms;

(3) administered the respective treatment (either conceptual description or practice opportunity); and

(4) displayed another query on the interface to each subject and then asked them to describe how they think the search engine would respond. Here, the approach from step 1 was reused to gain insight into the subjects’ mental models after the treatment. The overall purpose of the approach was to assist subjects,
without influencing their thinking, in expressing their understanding of how the system performs the query tasks.

In phase 2 we investigated subjects’ information seeking after receiving their respective treatment. During this phase we:

(1) asked the subjects to conduct a search task from the KITE case library. This information seeking task required a user to find a best match case that meets certain known characteristics. Specifically, the task asked the user to find a case about how a teacher uses “educational software” in lesson to motivate second-grade students to learn about dinosaurs. This task was developed and pilot-tested based on typical tasks teachers often use in the classroom. This type of information seeking task is frequently used in research studies in the field of library and information science (Allen, 1996; Hsieh-Yee, 1993; Palmquist & Kim, 2000). As a matter of fact, the case library does not contain a case that includes all of the criteria stated in the scenario. The search task session was recorded using a screen capture software program;
(2) interviewed each subject by asking them to indicate: how satisfied they are with the overall Search; how they would use the case they found for the task scenario; reflect upon their searching experience, review the search results, and then explain how the KITE search engine operates;
(3) evaluated the subject’s search by rating their search results; and
(4) asked the subjects open-ended questions about their attitudes towards the KITE and about their general impressions about the study.

7.3. Variables

In the phase 1 of this study, the independent variable is the two treatments. One is the conceptual description of the KITE CBR search engine. The other is search practice (system exposure) on the KITE CBR retrieval system. The dependent variable is users’ mental models.

During phase 2 of this study, the independent variable is users’ mental models. The dependent variables are search satisfaction, search correctness and search time, as follows:

- **Search satisfaction.** After the search task, subjects were asked to indicate how satisfied they were with the search results on a scale of 1–5 with 5 meaning “extremely” and 1 meaning “not at all.”
- **Search correctness.** In this study, the task was designed to find a specific case from the KITE case library. We determined by rating subjects’ answers on a scale of 0–2. If the best case was identified, the subject would get two points. One point was given if the subject found the relevant cases. For any other cases, the subject received zero points. The relevance of cases to the task was also verified in consultation with peers and participants during the pilot study.
- **Search time.** We measured the subjects’ search starting time and stopping time for the search task. The maximum search time allowed was 15 minutes. The reason for the time limit was to reduce the potential effect of fatigue. The pilot study found that 15 minutes was sufficient time to complete a search task.
- **Mental model.** In this study, mental models relied on qualitative data collection because we wanted to gain a rich understanding of how users perceive, explain, and cope with potentially confusing search engine operations. After the treatment, if there was a concrete evidence through the interview that the subject knew the search engine did a best or close match (match based on the relevance to the semantics closeness of the search terms, as opposed to an exact keyword match) then the subject was considered to have a “best match” model. The analysis began with a qualitative content analysis of the interview transcript. Then we examined the relevant evidences and made a decision whether the subject has clearly presented a “best match” model.

7.4. Data analysis

The data analysis began with a data screening procedure to check accuracy of data entry, missing data and outliers. As a result, 40 subjects were eligible for the phase 1’s mental model analysis. However, only 35 sub-
jects in total were eligible for the phase 2’s data analysis. In phase 2, there were 19 subjects in the search practice group \( n = 19 \) and 16 subjects in the conceptual description group \( n = 16 \).

The data collected were both quantitative and qualitative. We recorded the subjects’ responses on the questionnaire forms and on the audiotapes. Immediately after each data collection session, we transcribed the audio recordings of the interviews. The questionnaire data, interview data, and the search logs for each subject constituted one data set. Qualitative (open-ended interview questions) and quantitative (Likert scale ratings) data were treated differently. The analysis began with a qualitative content analysis of the interview transcript. We applied the deductive category application approach (Mayring, 2000) to analyze the transcript. During the coding process, mental models were classified into two categories: exact match and best match. Each subject’s prediction and explanation of the search engine’s operation was used as raw data to match against the two categories. For example, a subject’s interview data indicated several times that keywords were used for case retrieval and returned results need to include one of the words. In this case, the subject was considered to have an “exact match” model. On the contrary, when a user indicated that the search engine did a close match and found something related, the subject was considered to have a “best match” model. This deductive approach provided an efficient way to tie the assignment of the category to a passage of text in the transcript. Then, numerical data from the questionnaire and the search logs from each subject were used to measure search correctness, search time, and search satisfaction. After all the data were recorded, the researchers used a chi-square test and an independent sample \( t \)-test \( (p < 0.05) \) to compare the two treatment groups respectively. In addition, the researchers used the \( t \)-test \( (p < 0.05) \) to compare how subjects with best match models and subjects without best match models differed in their search correctness, search time, and search results satisfaction. The Pearson’s correlation analysis method was used to analyze the relationships among variables. The researchers also examined search logs to collect information about individual search actions and to measure time.

8. Findings

The findings of the study are summarized as below:

- \textbf{H1: There is significant difference between conceptual description and search practice treatments in terms of the effects on the change of the types of user's mental models.}

Table 1 lists the distribution of mental models after treatments in phase 1. Twelve subjects out of 20 (60%) in the conceptual description treatment group presented a best match model. Eight subjects out of 20 (40%) in the search practice treatment group presented a best match model (Table 1). We performed a chi square test in the SPSS program and found that there was no significant difference between the conceptual description treatment group and the search practice treatment group in terms of their effects on the change of the types of user’s mental models \( (p > 0.05) \). We also performed chi-square test to analyze whether there were any significant associations between the sample characteristics and the types of user’s mental models. There were no significant associations found between them. The same results also applied to the Table 2 which lists the distribution of mental models in phase 2.

During the post-test interview, when asked what the KITE search engine would do when choosing ‘simulations’ from Technologies used in Lesson and ‘music’ from the Subject/Unit, the subjects gave diverse responses. The following quotes are representative of the subjects’ responses in the search practice treatment group:

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Types of subjects’ mental models after phase 1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Post mental model</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>
‘‘I did the practical searches required for the study, and none of the tasks revealed the results I expected. It may be because the database did not have the information specified.’’

‘‘I believe that keywords are used and articles that may have one or both of the words included’’.

‘‘Search engine uses ‘‘OR’’ not ‘‘AND,’’ to produce query results.’’

‘‘I cannot even guarantee that the search engine would draw up anything related to simulations or music at all. It could bring up something totally unrelated.’’

Similarly, the researcher received diverse answers from the subjects in the conceptual description treatment group. Below are some representative quotes from that group of subjects:

‘‘It is an OR find. It would be articles with simulation or music.’’

‘‘It will do a similarity search and find something close or related such as...’’

‘‘It may find articles relating to simulation, music or just one of above categories. It may also find anything related too.’’

‘‘It will find articles including music & simulations, or science & simulations, or music & some other technologies.’’

‘‘The search will expand to other terms in the same category and do a broad search. It will find anything related.’’

‘‘An article would come up about music and the use of simulations. How they pertain to the other, such as instruments, music waves, etc.’’

Qualitative content analysis indicated that subjects in the search practice treatment group noticed a difference in their query results during their search practice opportunity. Therefore, some of the subjects did figure out that the search engine was searching meaning-related cases. They understood what the relevance percentage meant and could predict what results they might get when asked the post-test question. They did present somewhat the characteristics of the best match model. But none of the subjects could articulate how the KITE CBR search engine exactly worked. They were not exactly sure of the procedure about how the search engine did the relevance match and ranked the results. The following comment was echoed by some of the subjects:

‘‘The search engine was doing close match. But I do not know how. I can not read the mind of the engine.’’

On the contrary, subjects with best match models in the conceptual description treatment group seemed to have more complete models than subjects in the search practice treatment group. Ten out of the 12 subjects with best match models (83.3%) knew the KITE search engine was expanding the search by adding search terms with different weights within the same search field. The subjects with best match models were also more positive about the future of this type of search engine and considered the search engine to be helpful and useful. A subject with a best match model stated:

‘‘It took a little while, but I started to understand how the terms interacted with each other and brought about results. Once I understood, I appreciated the search engine more.’’

Finally, we noted that quite a few subjects in the two treatment groups were not able to come up with a plausible explanation or a mental model for the search results during the post-test in phase 1. They had no understanding about how the KITE search engine processes queries. They were surprised by the different results and did not appreciate the search results. Some of the subjects thought that the search engine was poorly designed because the search results were confusing and did not make much sense to them. They
complained that they could not find what they wanted. One subject described the surprise and confusion she had in her search practice:

“Although when searching for home economics subject and Microworlds technology, it did find Home Economics subject yet it ordered in a 5th place and the 4th place showed an article that did not meet such requirements.”

Instead, these subjects said they would like to have a keyword search function on the search interface, and they suggested a Google-type keyword highlight feature on the search result pages in order to help them identify the keywords quickly. Therefore, these subjects’ mental models were considered unchanged, remaining in the “exact match” type.

• **H2:** There is a significant difference between conceptual description and search practice treatments in terms of the effects on search correctness.

Subjects in the conceptual description treatment group scored a group mean of 1.31 (standard deviation = .704; range = 0, 2). Subjects in the search practice treatment group scored a group mean of 1.11 (standard deviation = .875; range = 0, 2). Statistics testing found that there was no significant difference between the two treatment groups in terms of their effects on search correctness \((t = 0.762, df = 33, p > 0.05)\). The lack of a significant difference in the two groups is understandable considering the lack of variability in the number of subjects with best match mental models in this study.

• **H3:** There is a significant difference between conceptual description and search practice treatments in terms of their effects on search time.

Subjects in the conceptual description treatment group scored a group mean of 339.06 seconds (standard deviation = 110.513; range = 132, 535). Subjects in the search practice treatment group scored a group mean of 249.74 seconds (standard deviation = 102.804; range = 68, 418). Statistics testing found that there was a statistically significant difference between the two treatment groups in terms of their effects on search time \((t = 2.475, df = 33, p < 0.05; Cohen’s d = 0.84)\). Specifically, the search practice group spent significantly less time than conceptual description group in finding the results.

• **H4:** There is a significant difference between conceptual description and search practice treatments in terms of their effects on search satisfaction.

The subjects in the conceptual description treatment group scored a group mean of 3.63 (standard deviation = .885; range = 2, 5). Subjects in the search practice treatment group scored a group mean of 3.63 (standard deviation = .831; range = 2, 5). Statistics testing found that there was no significant difference between the two treatment groups in terms of their effects on search satisfaction \((t = 0.023, df = 33, p > 0.05)\). Similar to search correctness, the lack of significant difference in the number of subjects with best match mental models contributes to the lack of significant difference in their search satisfaction. Small sample size may be another reason.

• **H5:** There are significant correlations among users’ mental models, search satisfaction, search time and search correctness.

The data screening procedures found that two subjects in the conceptual description treatment group were univariate outliers. Moreover, two subjects in the conceptual description treatment group and one subject in the search practice treatment group gave up their search tasks. Therefore, after five subjects were excluded, a total of 35 subjects were found eligible for phase 2. Specifically, 19 of these 35 subjects were in the search practice group \((n = 19)\) and 16 were in the conceptual description group \((n = 16)\). Nine subjects out of 16 (56.25%) in the conceptual description group had a best match model. Eight subjects out of 19 (42.11%) in the search practice group had a best match model.
We performed a chi square test in SPSS and found that there was still no difference between conceptual description and search practice in terms of the types of user’s mental models ($p > 0.05$). There were also no significant associations between the sample characteristics and the types of user’s mental models.

Next, we used the $t$-test method ($p < 0.05$) to compare how subjects with a best match model and subjects without a best match model differed in their search correctness, search time and search satisfaction.

As a result, subjects with a best match model scored a group mean of 4.18 (standard deviation = .529; range = 3, 5) on search satisfaction, a group mean of 318.82 seconds (standard deviation = 86.455; range = 170, 460) on search time, and a group mean of 1.47 (standard deviation = .717; range = 0, 2) on search correctness. Subjects with an exact match model scored a group mean of 3.11 (standard deviation = .758; range = 2, 4) on search satisfaction, a group mean of 263.89 seconds (standard deviation = 132.124; range = 68, 535) on search time, and a group mean of .94 (standard deviation = .802; range = 0, 2) on search correctness.

The statistical test in the phase 2 study found that subjects with the best match mental models have significantly higher search correctness ($t = 2.041; df = 33; p < 0.05; \text{Cohen’s } d = 0.7$) and search result satisfaction ($t = 4.794; df = 33; p < 0.05; \text{Cohen’s } d = 1.64$) than subjects without the best match mental models. However, best match mental models do not necessarily mean less search time in finding the results. It was surprising that we did not find significant correlation among search time, search correctness and search satisfaction. Further research may be needed to take a closer look at the measurement of these variables.

9. Discussion

The KITE CBR search engine was not designed to provide an exact match to the user’s query, but rather the closest match. Many subjects in this study were intrigued by the novelty of the KITE CBR search engine. The results of this study show the importance of training and mental models for the successful use of CBR retrieval systems. While efforts are being made to increase the usability of the search engine, progress in the area of CBR system interface design has been slow. Effective system training may be the key to increasing the users’ search success and decreasing user frustration while using this type of system.

The study found that both search practice and conceptual description have different effects in improving people’s use of the CBR systems. Subjects in the search practice group spent significantly less time than subjects in the conceptual description group in terms of the completion of the search task. However, subjects in the conceptual description group had more complete mental models than those in the search practice group when asked to explain how the CBR retrieval system processes the query. Furthermore, this study shows that subjects with the best match mental model showed higher search correctness and search satisfaction than subjects without the best match mental model. In contrast, most users without a mental model of the KITE case library expressed frustration while searching.

The results of the study addressed an important question: what is the effectiveness of two different methods of forming mental models among novices of a relative unexplored type of retrieval system, case-based reasoning retrieval system? The study also validated some previous findings in the new context of CBR regarding how mental models affect search performance and satisfaction. In addition, the study provided a new perspective on analyzing CBR system usability by using mental models as a framework.

The findings also provide practical insights for those who design and develop the CBR retrieval systems and interfaces. The CBR system and interface designers need to think about ways of preventing users from getting lost in searching the CBR systems. A practical suggestion for the CBR system and interface design is to make conceptual description clearly available on the interface. The conceptual description could be displayed on the interface in different ways. Besides putting a conceptual description on the search interface, the CBR system and interface designers could integrate a conceptual description into the users’ search experience with the system by providing conceptual feedback or explanations in the search result pages and suggesting similar cases as options when there are no exact match results found. Another suggestion we can make is to design a learning environment which can be used as a training and practice facility for the users of the CBR retrieval systems. The learning environment should provide users the opportunity to practice, learn and possess conceptual schemas for using the CBR retrieval system. In order to facilitate the
development of users’ mental models of the CBR retrieval mechanism, we suggest that the learning environment highlights the similarities and differences with other information search environment such as Boolean-based databases.

This study has several limitations. First, subjects in this study represent a convenience sample. They were selected through their self-reported membership and self-reported experience with computers, Web searching, instructional technology and case-based reasoning systems. The study sample size is also relatively small, although a substantial amount of time and effort was spent to gather a suitable sample. Although this was not particularly problematic for the study, it did mean that the results cannot be generalized to the population as a whole. Second, the KITE super search engine is not a representative of all case-based reasoning retrieval systems. The KITE super search interface and function also cannot be applied to all best match systems. Therefore, the findings of this study can only be applied to systems that are similar to the KITE system. Finally, as mental models are the internal conceptual representations that people develop while interacting with the environment, people normally have difficulty articulating the structure and content of their own knowledge of the systems they interact with (Briggs, 1987; Norman, 1983; Zhang & Chignell, 2001). Therefore, it should be noted that the mental model measured in this study was an external representation of the user’s understandings of the CBR search engines.

10. Conclusion and future research

Case-based reasoning systems are increasingly common on the Web. However, little empirical research has been done to systematically investigate people’s mental models in using this type of system. CBR research traditionally focused on the theories and heuristics of case representation, retrieval, reuse, revision, and retention. Little attention was paid to the user study of CBR systems such as user training, users’ mental models, and interaction with the systems. We strongly believe many CBR systems would fail to gain wider acceptance without adequate consideration to the interaction between the system and its user. In response to needs for the CBR user study, this study investigated the effects of conceptual description and search practice treatments on users’ construction of mental models, search time, search correctness and search satisfaction in a case library with a CBR retrieval mechanism. The study also sought to understand how the presence of a mental model affects the user’s search performance and satisfaction in this case library.

In conclusion, as an exploratory research at the intersection of information science and learning technology, this research revealed the effects of the conceptual description and search practice treatment on the users’ understanding and information seeking of best match systems. The research findings make a contribution to empirical knowledge on how treatments can be used to help users construct mental models in CBR retrieval systems. The knowledge will help system designers and trainers to create better user support mechanisms that will assist users in forming appropriate models for successful utilization of a new generation of information retrieval systems. Furthermore, the study revealed how the presence of a best match mental model affects the users’ search performance and search satisfaction in a CBR retrieval system. Prior to this study, it was not clear how users’ mental models affect their interaction with the case-based reasoning retrieval systems.

Regarding future research, additional populations of users need to be studied. In this study all the subjects were undergraduate students from a social and human science background. As there may be other user characteristics that influence users’ mental models, studies of different populations need to be conducted to help understand the mental models and information seeking in best match systems among users in different fields of study or some other research parameters. A large-scale study with different task complexity and difficulty is also needed to determine the effects of tasks on mental models. More research is needed to gain an understanding of users’ mental models over time and to explore how users employ their mental models with information seeking in various tasks and contexts. Particularly, we suggest that further research is needed to take a closer look at specific points of difficulty during each user’s search process and explore how the difficulty impacts or reflects the change of the user’s mental model conditions. This understanding will help system designers to design best match systems that better support users by allowing them to overcome the search difficulty and easily adapt their mental models from exact match to best match.
Acknowledgement

Many thanks to the anonymous reviewers, for their valuable suggestions to improve the content and the structure of the paper. This project was supported by the US Department of Education under grant number P342B010016.

References


**Wu He**, Ph.D., is an Instructional Technologist and Assistant Professor in the Center for Learning Technologies at Old Dominion University in Norfolk, Virginia.

**Sanda Erdelez**, Ph.D., is an Associate Professor of the School of Information Science and Learning Technologies, University of Missouri-Columbia.

**Feng-Kwei Wang**, Ph.D., is the Executive Director of ITRI College, Industrial Technology Research Institute (ITRI), Taiwan.

**Chi-Ren Shyu**, Ph.D., is an Associate Professor of the Department of Computer Science, University of Missouri-Columbia.