A Systematic Understanding of Successful Web Searches in Information-based Tasks

Mingming Zhou
National Institute of Education, Nanyang Technological University, Singapore // mingming.zhou@nie.edu.sg

(Submitted August 23, 2011; Revised September 27, 2011; Accepted January 13, 2012)

ABSTRACT
The purpose of this study is to research how Chinese university students solve information-based problems. With the Search Performance Index as the measure of search success, participants were divided into high, medium and low-performing groups. Based on their web search logs, these three groups were compared along five dimensions of the search process for answers to the assigned tasks: frequency counts of different activities during search process, time allocation on each search activity, search patterns, search query as well as selection of information. The results documented students’ varying abilities to search. The key factor that determines search performance is the effective use of search queries. Successful searchers made use of advanced search options such as extracting relevant and effective key terms from further reading, in contrast with simpler strategies by unsuccessful searchers like adding, removing unimportant words, or using synonyms. High-performing students also showed more monitoring awareness and strategy, such as changing search engines, reformulation of existing search queries as well as backtracking to task questions.

Keywords
Information problem solving, Log analysis, Web search skills

Introduction
In the information era, both social and technological developments have contributed to a situation where information plays a critical role. Computer searching has consequently become the routine behavior for identifying and assessing information to build knowledge. Based on comScore figures, 213 million searches are performed per day within the United States in March 2006. Thus, it is increasingly important that students master the skills to locate, select, evaluate and synthesize information from various sources to meet individual needs especially in their academic studies. Together, these activities constitute a process referred to as information-problem solving (IPS) (Brand-Gruwel, Wopereis, & Vermetten, 2005). This process has been extensively studied, and has resulted in a variety of descriptive and prescriptive models supported by empirical evidence (e.g., Brand-Gruwel, Wopereis, & Walraven, 2009; Ellis, Cox, & Hall, 1993; Kuhlthau, 1993). Across these models, typical elements involved in IPS include identifying learning needs/task interpretation, locating information sources, extracting and organizing relevant information, and synthesizing information from selected sources into cogent, productive uses. In this study, I examined how Chinese university students solve information-based problems and whether the process differs as a function of their performance levels, using computer logs generated while the students performed the task.

Related work
A brief overview of recent literature on IPS clearly reveals that researchers have shifted from investigating search product to search process. Of particular interest to researchers is the comparison between experts and novices in terms of their search strategy use and outcome. Most of these studies have established that experts are more successful in carrying out search tasks than novices. More specifically, experts typically tend to (a) use longer queries (Aula, 2003; Hölscher & Strube, 2000), (b) plan ahead in their searching behavior (Navarro-Prieto, Scaife, & Rogers, 1999), (c) make greater use of Boolean operators and modifiers (Aula & Siirtola, 2005), (d) use more keywords and evaluate the sites based on established criteria (Tabatabai & Luconi, 1998), and (e) pay more attention to the analysis of problems and the evaluation of solutions (Land & Green, 2000). However, in some studies, only small differences were found between the experts’ and novices’ search strategies (Brand-Gruwel et al., 2005) and outcomes (Navarro-Prieto et al., 1999). Another line of research compares search behavior and outcome among domain expert and non-expert. Research consistently shows that domain experts had superior performance over domain novices in terms of efficiency and effectiveness. That is, experts took less time to complete the search tasks and produced a larger number of correct solutions (Lazonder, Biemans, & Worpeis, 2000). They adopted longer and
more complex queries than novices, and used elaborations as a reformulation strategy more often compared with the simple stemming and backtracking modifications used by novices (e.g., Hembrooke, Gay, & Granka, 2005; Zhang, Anghelescu, & Yuan, 2005).

Nonetheless, variation exists in how the level of expertise or web experience is distinguished. Some studies propose that users with a little over one working week (50 hours) of experience are experts (Lazonder et al., 2000); others view final-year doctoral students (Brand-Gruwel et al., 2005), or Internet professionals as experts (Hölscher & Strube, 2000). The label “novice” has been assigned either to individuals with less than five years of computer experience and less than one year of web experience (Jenkins, Corritore, & Wiedenbeck, 2003), psychology freshmen (Brand-Gruwel et al., 2005), or psychology students with one year of web experience (Navarro-Prieto et al., 1999). Thus, the experts in one study could easily be novices in another. The same applies to the definition of domain experts and non-experts. The measurement of subject-domain expertise ranges from having specialized training or having performed research in the field (Jenkins et al., 2003), the achievement score in the subject (Lazonder et al., 2000), answers to fact-based questions (Duggan & Payne, 2008), to self-rated familiarity with the domain (Hembrooke et al., 2005). Aula and Nordhausen (2006) argued that this ambiguity in definition has resulted in difficulties in comparing the results from different studies directly, which does not add much to the understanding of the characteristics of different groups. Also the observed contradictory results from different studies may not actually be contradictory, but simply due to the fact that very different groups of users are being compared.

In addition to the terminology related to search experience or domain expertise, the way the search outcome is measured also varies considerably between studies (Aula & Nordhausen, 2006). Different measures for success are used, such as task completion time, number of tasks solved successfully, number of pages browsed, or accuracy of the answers, which again makes it difficult to compare the results across studies. One shortcoming in using task completion rate alone as a measure of search success is that a searcher can be very slow but eventually find the correct answer, whereas another could find the answer almost immediately yet the quality of the answer is much poorer. As both effectiveness and efficiency are important in determining the success of search, they need to be considered together, not independently as in most of the previous studies.

To address this issue, Aula and Nordhausen (2006) employed the level of experience as one variable predicting search success, which was measured by the task completion time (efficiency) and the number of tasks completed (effectiveness), which were combined together to form a single measure – task completion speed (TCS), to avoid the risk of considering fast but unsuccessful performance as superior to slow but successful. Using this approach, they found that the increase in the years of web use, speed of composing queries, average number of query terms per query and the participants’ evaluation of their search skills were all related to improvement in TCS.

Although Aula and Nordhausen’s (2006) study provided valuable insight into the successful strategies in IPS beyond the previous knowledge of the expert–novice differences, the data was retrieved from observations, think-aloud, and post hoc interview. These methods indeed provide invaluable data concerning the mental activities involved in the process, including the reasons, intentions for the engagement with certain behavior. Yet they are useful in uncovering new or emerging phenomena, rather than testing or confirming what is already known (Young, 2005). Further, given that individuals have a limited short-term memory capacity for talking and attending at the same time, the process of thinking aloud may consume limited attentional resources (Schraw, 2000), and their cognition may differ from what would have been the case had they not verbalized by thinking aloud (Leow, 2002). Thus, task performance of participants who used this method could decline (van den Haak, de Jong, & Schellens, 2003).

**Tracing method in IPS research**

Winne (1982) introduced the term *traces* to describe data that reflect actual learners’ cognitive engagements in learning tasks. Traces are accretion data gathered relatively unobtrusively as learners work on tasks. In contrast to traditional measurement, computer traces can document the dynamic, situated nature of learning, as well as individual event-based differences in activity (MacGregor, 1999). Computer traces are believed to be more accurate than self-reports that require recall of actions and advantageous over concurrent think-aloud protocols, which have been found sometimes to interfere with task performance (van den Haak et al., 2003), and live researcher observations which are constrained by human ability to attend to and record detail (Nesbit & Hadwin, 2006). Empirical data further proved that there is little or no correspondence between the retrospective self-reporting on one
hand and the online log-file measure of students’ cognitive processing on the other (e.g., Sins, van Jooying, Savelsbergh, & van Hout-Wolters; 2007; Winne & Jamieson-Noel, 2002).

With computer data, information such as search queries, hyperlink URLs, timing of each action and so forth can all be available. The frequency counts, timing and patterns of different study actions can be easily identified and, together with other learner characteristics, contribute to explaining the level of task performance. More importantly, the regulation of behavior during IPS process is made explicitly and available in this approach with high accuracy guaranteed. Computer screen capture tools have been found quite useful which are used to record searchers’ mouse behavior as well as audio input in real time (e.g., Buhi, Daley, Fuhrmann, & Smith, 2009; Tu, Shiha, & Tsai, 2008). In this way both students’ search behaviors and the reasons behind their behaviors could be analyzed and understood. Yet the main data analysis method is restricted to labor intensive video data analysis. Another type of computer data collection for online learning is through log files (e.g., Ford, Miller, & Moss, 2005). This format of data allows the use of advanced analysis techniques that could have examined the research questions from a different perspective. While this unobtrusive and accurate record (i.e., computer videos and logs) offers greater details about how students search, it only serves well the analysis of actual search behavior as students’ implicit thoughts and beliefs during the search are not necessarily transferred to explicitly observable online behaviors.

The use of traces in the investigations of students’ online search behavior has gained more and more attention to describe and analyze this dynamic process. According to Jansen and Spink (2006), the studies in this area fall into three categories: (1) those that describe users’ search behavior in an experimental setting, (2) those that use transaction log analysis over a period of time or across a number of users, and (3) those that examine issues related to or affecting web search. This paper focuses on the first line of research. Unfortunately, only a few empirical studies in IPS research up to date fall into this category. Roy and Chi (2003) used computer logs, coupled with field notes, to explore how eighth-grade boys versus girls searched on the web for a task. Each student’s search behavior was diagramed out and a series of six different “search moves” were derived. Results revealed that boys tended to employ a different search pattern from girls and that this variation in search behavior was related to the pattern of performance outcomes. In Leroy, Lally, and Chen’s study (2003), the participants were divided into three groups according to overall actual performance and showed how each group interacted differently with the query optimization feedback. Log analyses showed that this feedback significantly helped low achievers but seemed to hinder high achievers. Again, these studies do not meet the goals of the current research.

In sum, there are studies on skilled versus unskilled searchers’ search strategies, studies adopting more reliable measures of search success, as well as studies using computer logs. However, studies focusing on all these aspects are scarce. The present study aims to meet this gap by using computer logs to investigate how adult learners engage in and regulate their IPS process and whether successful searchers exhibit different search patterns from unsuccessful searchers. The underlying questions of this research are: (1) what types of activities students engage in when solving information problems, (2) how students regulate their IPS process, and (3) whether successful searchers show a different search pattern from unsuccessful searchers.

Method

Participants

The participants were recruited via advertisements in a Chinese university. Twelve male Chinese university students majoring in computer-related areas (e.g., software engineering, telecommunication, etc.) volunteered to participate in this study (mean age = 20.3). On average, they spent 37.7 hours per week on the Internet. Among different online activities to accomplish academic tasks, searching information was reported to be the most frequent activity (rating of 4 out of 5), followed by downloading relevant material from course website (rating of 3.3 out of 5), and searching materials in the university’s online library (rating of 3 out of 5).

Apparatus

During the search tasks, the participants used a PC workstation. All participants used Firefox as the web browser. Google was set up as the default search engine, but the participants were allowed to use any search engine they preferred.
preferred. All searches were performed in Chinese. For the data analysis, a computer tracking tool, SCOOP (Zhou, Xu, Su, & Liu, 2011) was developed which supports web search and collects online behavioral data (traces) without inserting potentially disruptive manipulations into student IPS. The tool records fine-grained traces of user click behaviors (time, duration, and object of each mouse-clicking), students’ navigation paths across the Internet as well as the content students input during the search (e.g., search strings). These actions traced during the student’s IPS session are logged to the second and analyzed to make inferences about student IPS activities.

**Tasks**

Two search tasks were assigned to the participants. The searching topic, how bees make decisions for new home locations, was designed in the way that (1) it was similar to academic tasks students usually encountered in university; (2) participants' prior knowledge would not vary much (the participants generally reported a lack of prior knowledge on the topic); and (3) the questions were sufficiently complicated to avoid successful search by only a few simple attempts (piloted with other undergraduate students with similar backgrounds). The task for the participants was to answer two short answer questions (200-300 words each):

1. How do bees choose where to build their new homes? (close-ended);
2. What do you think are the implications for human life? (open-ended).

Both the task questions and students’ answers were presented in Chinese. The students were allowed to finish the task in their own paces. They were also encouraged to highlight the information they found relevant to the task by using the highlighter in SCOOP.

**Procedure**

The study was conducted in a computer laboratory at the university. After arriving at the laboratory, the researcher informed the participants about the purpose and the procedure of the study. The participants were told that the purpose was “to study their normal information search strategies,” without knowing that their search behavior was tracked at the same time they search. As the aim was to make the search session as typical as possible, it was also emphasized that the students could use any search engines as they wished to complete the tasks. There was no strict time limit for the individual tasks. However, the time limit for the whole search session was limited to one hour. The participants were shown how to log in and make highlights before they start the task. To collect the information they found, the participants were encouraged to highlight the information they found relevant for the task.

**Data analysis**

Four students’ logs were lost or incomplete due to their failure to follow the experiment procedure. Hence, 8 logs constituted the dataset for the analysis below.

**Search performance index (SPI)**

Search Performance Index (SPI) is a measure of the performance in the search tasks, which considers both the efficiency and effectiveness of the search, both the accuracy of the answers (effectiveness) and the total time of task completion (efficiency). The SPI is similar to the task completion speed used by Aula and Nordhausen (2006). However, in their measure, the effectiveness of the search is indicated by the number of tasks completed, whereas the actual score of the answer to each task is employed in this study. Given there are only two tasks in the current study, the use of scores of the answers provides greater variation when interpreting the difference among participants. The SPI measure was calculated with the following formula:

$$SPI = \frac{Accuracy}{Time} \times 60$$

In this formula, total time of task completion is entered in minutes. Thus, the quotient (accuracy of answers/task completion time for both tasks) is multiplied by 60 to denote the speed of the task accuracy per hour. The accuracy
of answers was scored between 0 and 5. For the first close-ended question, participants received 5 points if the answer is to the point (i.e., answering what has been asked in the question), elaborative (i.e., including one or more supporting statements or evidences), and logic (i.e., explaining the ways bees select new home locations logically). For the second question, students needed to search web resources, analyze and critically evaluate web materials, and possibly put personal thoughts into answers to complete that task. There were no standardized answers for the second question. The students’ responses of this question were rated on their soundness, richness and organization from 0 to 5 points. In sum, the total score of the two searching tasks were 0–10. Each participant’s answer was reviewed by two faculty members. The final score was the average of the two. Any difference more than three points were resolved through discussions.

The participants were split into three levels based on the distribution of their SPI (M = 10.87, SD = 11.9): high performance (N = 3; M =23.47) medium performance (N = 2; M = 8.29) and low performance (N = 3; M = 0). The three groups were compared along five dimensions: frequency counts of different activities during search process, time allocation on each search activity, search patterns, search query as well as selection of information.

Computer log analysis

Leard and Hadwin (2001) identified four major categories of log file analysis employed in the literature: (a) frequency analysis, (b) patterns of activity, (c) time-based analysis, and (d) content analysis. This guides the analysis of log files in this study.

Frequency analysis, the most prevalent method, involves recording the frequency of specified learner actions, such as accessing a relevant web page. Frequencies are usually treated as continuous variables and analyzed with parametric methods (e.g., analysis of variance or regression analysis). A limitation of frequency data is that it describes actions at one point in time but does not capture relationships across events or the time invested in those event-based actions (Misanchuk & Schwier, 1992). Patterns of activity address that limitation by considering contingent frequencies of specific actions or events. When activities occur in consistent proximity to one another (in a particular sequence), they provide information about learners’ strategic actions. Two common approaches for analyzing patterns of activity are transition matrices and time-based diagrams. The former is most commonly used to track two-event sequences, whereas the latter provides a visual overview of a learner’s entire interaction with the learning environment with each event types represented by symbols graphed on a timeline.

Great potential exists for examining not only the occurrence of events in sequence, but also duration and overlap in those events. The timing of every event available from the logs can be used to infer the duration of time students spend on particular activities, and total or mean time spent studying overall. Inferring duration from log data is important because it cues the researcher to consider what is going on for the learner between logged events and to consider time on task as an important element of engagement (Rouet & Passerault, 1999). To date, techniques for incorporating duration into the graphing of activity patterns have not been adequately explored (Nesbit & Hadwin, 2006). Meanwhile, it becomes important to examine the actual content students develop or work with to evaluate the depth of cognitive processing, to the extent that learning environments afford opportunities to create information objects such as highlighting, notes, etc. In IPS research, this analysis technique is most relevant to the choice of search queries.

Results

Frequency analysis

Frequency counts were calculated for each group on the main searching activities as shown in Figure 1. Consistently, the medium-performing group was more active in most aspects of search process. They visited and revisited more web pages, reviewed the task questions more often and made more highlights than the other two groups. The high-performing group used more search queries, and switched the search engine most frequently. Clearly, attempts with different search queries and search engines were deemed as effective strategies to locate information. The low-performing group consistently scored lowest in most dimensions. They accessed much smaller amount of information, did not review the task questions as much as the other groups and made least highlights during search.
The lack of sufficient engagement with the task could account for their inactiveness to some extent. Interestingly, this group used the same number of search queries on average as the high-performing group and tended to revisit the web pages more often than the high-performing group.

Time-based analysis

The frequency counts of learner activities provide overall information for a given individual or group by taking the problem-solving process as a whole, but they fail to reveal the total or mean time spent overall or on each component of the search process as a measurement of engagement. This can be solved by time-based analysis. As shown in Figure 2, the medium-performing students spent most time in total, followed by the high-performing and low-performing group. When the total time was broken down, the high-performing group typically allocated more time on searching (including constructing/updating search queries and evaluating web pages from the hit list), whereas the medium-performing group scored highest in reading the content of selected web pages and answering the questions. Surprisingly, they spent least time in searching. The low-performing group consistently spent less time on all stages of search. Within each group, both medium- and low-performing groups spent most time in constructing the answers, followed by reading online information then searching. Yet the high-performing group allocated slightly more time to accessing information than task writing.

Patterns of activity

Prior research (Branch, 2001; Brand-Gruwel, et al., 2005) repeatedly shows that students lack regulatory skills to successfully solve information-based problems. In this study, there was no instance wherein students located an answer with only one attempt. In other words, students needed to adapt their search strategies when they were unsuccessful during their initial attempts. Based on the log files, five major activities were observed during IPS:
search (submit a search query), read information (open and view a web page), review task instructions (click the web page which contains the task questions), highlight relevant information, and write answers. To better capture the regulatory process, pattern-based analysis was conducted to describe the sequence of activities performed during a specific period of time and presented in a time-based diagram. The analysis only involved the high and low performance group in order to elucidate the difference between successful and unsuccessful searchers to the greatest extent. A time-series sequence chart was created for each participant from these two groups. Levenshtein distance was employed to measure the similarity between pairs of sequences within the same group to search for a most representative pattern in each group. It is defined as the minimum number of edits needed to transform one sequence into the other, with the allowable edit operations being insertion, deletion, or substitution of an event in the sequence (Kruskal, 1983). A sequence with the lowest mean distance with all other sequences compared to within its group will be taken as the “most representative” pattern for that group. This pattern recognition strategy is not ideal in that it does not capture most commonalities across sequences in the same pattern. Yet it serves the purpose well of selecting a sequence that resembles all other sequences in the same group to the largest extent, especially with a rather small group size. In this way, subject 01 (SPI = 24) was identified to represent the high-performing group (Figure 3) and subject 04 (SPI = 0) for the low-performing group (Figure 4).

The two representative patterns differed in several ways. The most striking difference was that the low-performing group started answering the questions much earlier than their counterparts. It could be interpreted in two ways: They found (or they believed they found) the answers in the first few attempts; or they simply wanted to complete the task as soon as possible. Based on their task performance, the first could be due to the incapability of evaluating the relevancy of information correctly, and the second could be the lack of motivation to perform the task as well as they could. Further, different patterns were observed during various stages of IPS. The high-performing subject focused on information-seeking in the initial stage. In order to find the relevant information quickly and accurately, he switched between “search,” “read information” and “access task questions” quite frequently. This can be taken as a monitoring process, as referring back to the task question to fine-tune the search string was an effective way to stay on task. This task analysis action was deemed as a very important facet in self-regulated learning models (e.g.,
Winne & Hadwin, 1998). In this stage, students generate their own perceptions about what the studying task is, and what resources are in place such that they can construct a strategy. The enactment of the strategy occurred in the middle stage in this group, as seen through their construction of the answers by accessing online information. In the final stage, this group chose to highlight relevant information and revisited their answers with a few more new searches. This can be seen as a final check of their answer. This review exercise exhibited their meta-cognitive control. In contrast, the low-performing subject seemed to have experienced a few unsuccessful search attempts at first, as reflected by their frequent revisit of the task questions. Until they found relevant information (they believed), they started working on the answers with no more search in between. The whole process was ended by a few other reading and highlighting, which implied the uncertainty of the subject about the accuracy of the answer.

Content analysis

Another common strategy observed in all participants was to try new search queries. Interestingly, the three groups determined to use a new search string in different ways. The high-performing group on average decided to use an alternative search query on average every 2.04 minutes, the medium-performing group every 1.04 minutes and the low-performing group every 0.80 minutes. It appeared that high-performing searchers would like to access more information from the hit list before rushing to call it a failure and try different search queries. To find out what search strings students used to better understand the reasons for their (in)effectiveness, content analysis was conducted on students’ construction of search queries as well as the information they considered useful. 7 students simply submitted the task questions in their original form to the search engine. The low-performing group generally used this approach in the first attempt, while the high-performing group used it after several unsuccessful attempts. Another common search approach among most participants was to search with multiple keywords extracted from the task question, such as “bees build new homes,” “bees choose location,” and “bees home implications.” The high performing group also tried phrases taken from the web pages they accessed from previous searches: “choice of bee homes → bees collaboration → formation of bee hives → where do bees build homes usually → bee colony → swarm intelligence” (subject 02; SPI = 14.63). These phrases were extensions of the task questions that revealed the focus of potential answers, such as “bee colony” and “swarm intelligence.” The low-performing group adopted simple strategies to adapt the query by adding, removing, or replacing words from the task questions: “bees choose new home location → bees build new homes → bees choose new homes → bees’ choice” (subject 04; SPI = 0). Ineffective search strings were also found in this group, such as “how bees build hives,” and “bees’ life style.”

Students were also asked to highlight information they deemed useful for the task so that researchers can further assess their information synthesis capability. Content analyses of their answers showed that students highly relied on the information they highlighted to construct their answers, yet the high-performing group did bring in their prior knowledge or paraphrase the information they accessed into a logic and meaningful answer. The organization among different bits of extracted information has been observed to be a problem in the low-performing group, such as the lack of transitions, unclear or illogical relationships between sentences, and inappropriate structures. This reflects their lack of capability to synthesize information effectively.

Discussion and conclusion

In this study, Chinese university students’ IPS skills were examined in terms of frequency counts of different activities during search, time allocation on each search activity, search patterns, search query as well as selection of information. The approach of this study agrees with the work of Aula and Nordhausen (2006), which focuses on the level of actual performance in search tasks as an indicator of expertise. SPI can be a feasible and reliable measure for search success, and it is best used when (1) the study consists of one or more tasks; (2) the score of search performance is continuous; and (3) the nature of multiple tasks is similar to be combined into one single measure. In this way, it can be ensured that the observed effective search strategies can be the strategies experts adopt.

In a nutshell, this study documented students’ varying abilities to complete a search task through searching information online. Some successfully accomplished the task, whereas some completely failed. This was largely associated with the different strategies adopted during the process. Consistent with Hölscher and Strube (2000), successful searchers showed more complex behavior upon unsuccessful attempts, such as reformulations of existing queries, changing search engines, as well as backtracking to task questions. The insufficient engagement in the above
activities has led to poor search performance. Interestingly, full engagement with these activities did not result in the best performance either, as evidenced by the medium-performing group. It seemed that reading (and rereading) more information, re-accessing the task questions and making more highlights did help with the search outcome, yet they were not the key. There are three possible reasons. Given that there was far less pronounced difference for the number of search queries used among the three groups, the quality of the queries could affect the search success. Successful searchers made use of advanced search options such as extracting relevant and effective key terms from further reading, in contrast with simpler strategies by unsuccessful searchers like adding, removing unimportant words, or using synonyms. Another reason was the ability of evaluating the effectiveness of a search query. Clearly, high-performing students spent much more time on trying with different search queries. The judgment of the ineffectiveness of a search string stemmed from the glance of the title and synopsis of each hit in the search result page. This capability greatly enhanced the efficiency of searching process and outcome. The third reason was searchers’ monitoring awareness and strategy. The high-performing group tended to formulate a strategy (e.g., find the relevant information by reviewing the task questions), and adapt the strategy whenever necessary (e.g., do a final check by revisiting the answer with a few more new searches), while low-performing students were not skilled at monitoring the process and less active in adapting their search strategies. They tended to be eager to finish the task while being restricted to the limited amount of information.

The results are not generalizable considering the rather small sample size, uniform gender composition as well as the educational background (all from the computer field). Hence, the study is more exploratory, and serves as a preparation of a future larger-scale study. Also, attention needs to be paid to cross-cultural differences in the patterns of online search. For example, European users generated slightly more queries per session and viewed more retrieved pages per query than American users (Spink, Ozmutlu, Ozmutlu, & Jansen, 2002). Thus, the current findings may not apply to web searchers in other countries. Further, other data-gathering methods can be used, including eye-tracking (Cutrell & Guan, 2007), self-reports (Rieh, 2002), and observation notes (Roy & Chi, 2003). Each technique captures the process from a distinct perspective with benefits and losses. Traditional techniques are still valuable in that they provide capture the mental activities that are not detectable via traces. Log analyses are promising in that they allow researchers to dig into student actual behavior for any ground inferences to be made. Thus, these methods are not orthogonal, but supplementary to one another. And they are expected to be used jointly to paint a fuller picture of student IPS process.

Future work needs to be conducted with larger samples, multiple tasks of different natures or other data analytical techniques. With a different task, strategic students are expected to exhibit different approaches to solve the task. With a large sample, the behavioral traces can be submitted to more advanced statistical analysis, such as regression analysis, structural equation modeling, and other web-mining algorithms for analysis, such as sequential pattern mining techniques to discover information seeking patterns; associate rule mining techniques to find correlations among students’ search terms and the web pages they view; clustering and classification techniques to group students with similar characteristics or into pre-defined categories based on given individual variables or search action sequences. Analysis results will enable us to better assess student IPS skills, as well as identify their difficulties and challenges they experience. The next step is then to explore how educators can use the findings to improve students' acquisition of more efficient IPS strategies. This information will help educators and content developers to develop instructional environments that scaffold web search skills to improve the process and performance.

References


