Exploring search task difficulty reasons in different task types and user knowledge groups

Jingjing Liu [a,*], Chang Suk Kim [b], Caitlin Creel [c]

[a] School of Library and Information Science, University of South Carolina, 1501 Greene Street, Columbia, SC 29208, United States
[b] Department of Computer Science, Southern Connecticut State University, 501 Crescent Street, New Haven, CT 06515, United States
[c] Greenwood County Library System, 600 Main St S, Greenwood, SC 29646, United States

ARTICLE INFO

ABSTRACT

Search task difficulty has been attracting much research attention in recent years, mostly regarding its relationship with searchers' behaviors and the prediction of task difficulty from search behaviors. However, it remains unknown what makes searchers feel the difficulty. A study consisting of 48 undergraduate students was conducted to explore this question. Each participant was given 4 search tasks that were carefully designed following a task classification scheme. Questionnaires were used to elicit participants' ratings on task difficulty and why they gave those ratings. Based on the collected difficulty reasons, a coding scheme was developed, which covered various aspects of task, user, and user–task interaction. Difficulty reasons were then categorized following this scheme. Results showed that searchers reported some common reasons leading to task difficulty in different tasks, but most of the difficulty reasons varied across tasks. In addition, task difficulty had some common reasons between searchers with low and high levels of topic knowledge, although there were also differences in top task difficulty reasons between high and low knowledge users. These findings further our understanding of search task difficulty, the relationship between task difficulty and task type, and that between task difficulty and knowledge level. The findings can also be helpful with designing tasks for information search experiments, and have implications on search system design both in general and for personalization based on task type and searchers' knowledge.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Online information searching has become an everyday activity in the lives of people around the world. With today's search systems, people can often easily find the information they need. However, at times, people may have a hard time searching for information to fulfill a particular information need, for example, planning for a trip, buying a car, or researching the literature on a topic. These *difficult* tasks can lead to users not finding desired information and accordingly cause frustration and/or system switch behaviors. Search systems aimed at helping people locate information effectively, efficiently, and enjoyably should make it an important goal to reduce task difficulty, assist users with difficult tasks, and increase users' satisfaction.

Task difficulty has been attracting more and more research attention in the field of information retrieval (IR). Some researchers studied query performance and query difficulty from the language model perspective (e.g., Carmel, Yom-Tov,

---

Darlow, & Pelleg, 2006; Cronen-Townsend, Zhou, & Croft, 2002). Others studied task difficulty from the search task level. Kim (2006) suggests that difficulty is the task doer's perception of task complexity. Li and Belkin (2008) define task difficulty as a subjective perception assessed by task doers. Studies along this line have examined users' search behaviors in difficult vs. easy tasks (e.g., Aula, Khan, & Guan, 2010; Gwizdka, 2008), users' perception of task difficulty before and after working on the tasks (Liu, Liu, Yuan, & Belkin, 2011), and the prediction of task difficulty from search behaviors and/or other factors such as user characteristics (e.g., Arguello, 2014; Gwizdka, 2008; Liu, Belkin, & Cole, 2012).

Despite the above research directions of task difficulty, little has been known about why users feel certain tasks are difficult and what makes them feel this difficulty. Even though a system can predict, from monitoring the users' behaviors, that they are having difficulty in their search, the system cannot help users overcome the difficulty without a further understanding of what caused the difficulty. This calls for an exploration of search task difficulty reasons, and Liu and Kim (2013) started to address this issue and developed a task difficulty reason scheme based on a lab experiment with 32 participants searching for 4 tasks. It would be helpful to test this scheme with more participants. In addition, it would be interesting to explore if the difficulty reasons vary across task types and across users with different backgrounds. In order to better understand the nature of search task difficulty and eventually benefit search system design, the current research attempts to explore the following research questions:

1. What are the reasons users perceive a search task is difficult?
2. Do the task difficulty reasons vary across different tasks and task types?
3. Do the task difficulty reasons vary between searchers with different levels of task topic knowledge?

2. Literature review

2.1. Search task difficulty and user behaviors

Search task difficulty has attracted quite some research attention in IR, especially in recent years. There is not a consensus on a definition of task difficulty. Some researchers suggest that task difficulty is subjective. Kim (2006) defined that difficulty is the task doer's perception of task complexity. Li and Belkin (2008) noted in their comprehensive task classification scheme that task difficulty can only be subjective, as assessed by task doers. Some other researchers define task difficulty based on certain objective criteria and/or measurements. For example, Aula et al. (2010) employed a user's task answer being correct or not as the criterion of the task being easy or difficult. Liu, Liu, Cole, Belkin, and Zhang (2012) designed tasks with different difficulty levels based on the retrieved results' precision@10 using the topic terms as search keywords.

One line of studies explored the relationship between task difficulty and searchers' behaviors. Kim (2006) examined the effects of task difficulty on user behaviors in three types of tasks: factual, interpretive, and exploratory. Task difficulty was elicited through users' ratings on a Likert scale based on their subjective judgments. It was found that in factual tasks, post-task difficulty was significantly associated with task completion time, and the numbers of queries and documents viewed; in exploratory tasks, user behaviors were significantly correlated with pre-task difficulty; but in interpretive tasks, most correlations between behaviors and task difficulty were not significant. Liu, Gwizdka, Liu, and Belkin (2010) also used users' subjective perceptions on task difficulty as elicited by their ratings on task difficulty questions. They examined through a lab experiment how user behaviors vary in tasks with different difficulty levels, as well as different types. They found that in difficult tasks, users had longer task completion time, issued more queries, viewed more content pages, and had longer dwell time on content pages.

Aula et al.'s (2010) took a more objective way to define task difficulty. It was determined by user success or failure in finding the answers to their task questions, which were closed information tasks that had a single, unambiguous answer. The authors conducted two studies to research task difficulty, one being a lab experiment and the other a large-scale study. Results showed that in difficult tasks, users formulated more diverse queries, used advanced operators more, and spent longer time on the search result pages.

Another line of studies has been attempting to predict search task difficulty from searchers' behaviors. Gwizdka and Spence (2006) examined how searchers' behaviors could indicate the difficulty of a factual information-seeking task. Task difficulty was self-assessed by users after each task. Their results indicated that higher search effort, lower navigational speed, and lower search efficiency were good predictors of task difficulty tested by regression models. In Liu, Gwizdka, et al. (2010) study, tasks were categorized as easy and difficult based on searchers' post-task judgment on tasks' difficulty levels. Searchers' behaviors were grouped at two levels depending on the time point when the behavioral factor value can be captured: the whole-task-session level factors whose values can only be obtained after a search session is done, and the within-task-session level factors whose values can be obtained while a search session is ongoing. They found that both whole-session level and within-session level user behaviors could serve as task difficulty predictors in logistic regression models. Whole-session level variables showed higher prediction accuracy, but do not enable real-time prediction. On the other hand, while within-session level factors can ensure real-time prediction, the prediction accuracy in general was mediocre, especially in certain types of tasks, possibly because of the limited number of within-session factors that were considered and used in their model.

Liu, Liu, et al. (2012) investigated users' behavioral differences between difficult and easy search tasks and built prediction models based on users' behavioral factors. The behaviors they used were at three levels, distinguished by the time point

when the measurements can be done: first-round level at the beginning of the search, accumulated level during the search, and whole-session level by the end of the search. They found that a number of user behaviors at all three levels showed differences between easy and difficult tasks. Models predicting task difficulty at all three levels were developed and evaluated. A real-time model incorporating first-round and accumulated levels of behaviors obtained fairly good prediction performance that was comparable with the model using the whole-session level behaviors that are not real-time. A similar study conducted by Arguello (2014) also attempted to predict task difficulty using logistic regression models. The study also considered first-round and whole-session level search behaviors and collected difficulty ratings in post-search questionnaires. Other factors considered in the study included searchers' interest in and knowledge of search topics. The study found that whole-session prediction was more effective than first-round prediction. The best approach for either level prediction was to combine a wide-range of features, and for different level prediction, the most predictive features were different. The study also found that users' level of interest and prior knowledge did not help with prediction performance.

Liu et al. (2011) examined how users' perception of task difficulty changes before and after searching for information to solve tasks using two user studies which employed carefully designed tasks of different types along several dimensions: task structure (subtasks being dependent upon or parallel with each other), task goal in quality (being specific or amorphous), and naming (being named or unnamed). It was found that while in some types of tasks, users' perceptions of task difficulty did not change before and after working on the tasks, in others, they did, either increasing or decreasing. It was also found that users' background factors, e.g., previous topic knowledge and search experience, do not necessarily correlate with their perceived task difficulty, or perceived difficulty change.

2.2. Task type in information retrieval

Much attention has been drawn on examining the effects of different tasks on information searchers' behaviors and performance. The basis of studying task type is to classify user tasks into different types along some task feature(s), for example, open-ended vs. closed (Marchionini, 1989), and fact-finding vs. information gathering (Kellar, Watters, & Shepherd, 2007). Li and Belkin (2008) constructed a comprehensive classification scheme that includes a number of dimensions of task features/ attributes: task product, complexity, and difficulty, to name a few.

One frequently examined task feature is task complexity, which was related to, but not the same as task difficulty. Vakkari (1999) developed a model integrating task complexity and user actions. Byström and colleagues conducted a series of studies (Byström, 2002; Byström & Järvelin, 1995) examining the relationships between task complexity, information types, and information sources. They defined task complexity as the users' "a priori determinability of, or uncertainty about, task outcomes, process, and information requirements" (Byström & Järvelin, 1995, p. 194). They found that task complexity was related to information type and information source selection. As task complexity increased, users needed more sources of information, more domain information and more problem solving information; they were less likely to predict the types of information they needed; and they were more dependent upon experts to provide useful information. Li (2008) defined and measured task complexity by the number of activities in a work task, or the number of search sources needed in a search task. She found that objective task complexity affected many aspects of searching behavior, including: the number of search systems consulted, portals visited, web result pages and items viewed, users' interaction with library resources, query-related interactive behavior, success, satisfaction, and the total task completion time. Following the same definition of task complexity as Li (2008), Liu, Cole, et al. (2010) found that high complexity tasks led to longer task completion time, more pages visited, more sources, and more queries.

Task types have also been classified by other features and have been studied regarding their effects on users' information behaviors. Qiu (1993) found that when engaging in specific tasks rather than in general tasks, users tended to adopt more structured search patterns. Moreover, users preferred to use browsing features for completing general tasks, but analytical searching in specific tasks. Kim (2006) looked at three types of tasks: factual, interpretive, and exploratory tasks, and found that task type significantly influenced number of pages saved and the ratio of pages viewed to pages saved. Kellar et al. (2007) examined how users navigated and interacted with web browser across four types of tasks: fact-finding, information gathering, browsing, and transactions. They found the information gathering task to be the most complex: users spent more time completing it, viewed more pages, and used the web browser functions most heavily. Toms et al. (2007) examined how search was affected by tasks with different information goals (decision making, fact finding, and information gathering tasks) and task structures (having hierarchical or parallel concepts in the tasks). They found that in the hierarchical tasks, users performed fewer queries but took more time to process the search results than in the parallel tasks. In tasks with different goals, users did not show notable differences in the number of queries and time spent to formulate queries. Li (2008) studied tasks along the "task product" feature, in which tasks included intellectual and decision/solution types. She found that, compared with decision/solution tasks, intellectual tasks involved more IR systems consulted and result pages viewed, longer queries, and higher self-ratings on task success. Also along the "task product" feature to classify task types, Liu, Cole, et al. (2010) found that mixed product tasks, compared with factual tasks, led to longer task completion time, more pages visited, and more sources referred to. Liu, Cole, et al. (2010) also found that along the "task goal (quality)" feature, amorphous tasks, compared with specific tasks, led to more sources referred to.

2.3. Task topic knowledge in information retrieval

In the rich literature examining the effects of user knowledge on information searches, researchers have examined two different types of knowledge: (1) subject domain knowledge, which is a person's knowledge of a general subject or domain, for example, medicine, law, or psychology; and (2) task topic knowledge, which refers to the user's knowledge with the topic of his/her specific task at hand. Zhang, Liu, and Cole (2013) found different effects of domain knowledge and topic knowledge on users' behaviors. Topic knowledge was significantly correlated with performance measures for individual tasks, and domain knowledge was significantly correlated with performance at more general levels, over multiple task sessions. Because of the differences, the following reviewed these two types of knowledge respectively.

The majority of previous studies on knowledge have focused on domain knowledge. Hsieh-Yee (1993) found that domain knowledge affected search tactics. As compared to a familiar subject area, when users worked with a search task outside of their field, they used the thesaurus more for term suggestion, made more effort in preparing for the search, monitored the search more closely, included more synonyms, and tried more term combinations. Vakkari, Pennanen, and Serola (2003) found that medical major students began to use a wider and more specific vocabulary in their development of research proposals at the end of a 3-month seminar compared to the beginning. Sihvonen and Vakkari (2004) found that the number and type of terms selected from the thesaurus for expansion by domain experts in the medical area improved search effectiveness, whereas the use of the thesaurus by domain novices had no impact. Wildemuth (2004) found that low domain knowledge, before the users took a microbiology course, was associated with less efficient selection of concepts to include in the search and with more errors in the reformulation of search tactics. Zhang, Anghelescu, and Yuan (2005) found that a high-level knowledge group of users in the Heat and Thermodynamics engineering domain were found to have better performance (retrieved slightly more relevant documents), issued longer queries, and had more queries per task. Duggan and Payne (2008) found that knowledge of the music domain had little effect on search performance, but that of the football domain had much effect on search performance: it was positively correlated with search accuracy, and negatively correlated with time spent on web pages and mean query length. White, Dumais, and Teevan (2009) found that within their domain of expertise, experts search differently than non-experts in terms of the sites they visit, the query vocabulary they use, their patterns of search behavior, and their search success.

Meanwhile, some studies investigated the effects of search task topic knowledge in information searches. Topic knowledge is often measured by asking subjects about their familiarity with the task topic in research (Kelly & Cool, 2002). In this paper, topic knowledge and topic familiarity are used interchangeably.

Allen (1991) defined "topical knowledge" as specific factual knowledge of a topic. He found people with high topical knowledge used more search expressions than those with low knowledge. Kelly and Cool (2002) found that reading time tends to decrease and efficacy (measured by the ratio of the number of saved documents to the total number of viewed documents) increases with increasing search task topic familiarity. Hembrooke, Granka, Gay, and Liddy (2005) found that topic experts issued longer and more complex queries than topic novices. Experts also used elaborations as a reformulation strategy more often as compared to simple stemming and backtracking modifications used by novices. Liu and Belkin (2010) found the total dwell time that users spent on a retrieved document was positively correlated with document usefulness regardless of the user's topic knowledge level, but the first dwell time (the unobtrusive duration from opening a document to the user leaving it for the first time) was not significantly correlated with document usefulness, and topic knowledge affected their correlation.

3. Method

3.1. Experimental design

To answer the research questions, a user experiment was conducted. A total of 48 college students participated in the experiment, coming from two universities, 16 at the University of South Carolina (USC) Columbia campus and 32 at the Southern Connecticut State University (SCSU) campus. The experiment asked participants to work with four assigned information search tasks, and provide their subjective judgments of task difficulty and reasons. Task type was the pre-controlled factor, with different features along various dimensions in a task classification scheme (details below). Task difficulty was obtained post hoc through user ratings on a Likert scale, as was users' topic knowledge levels.

3.2. Participants

Among the 48 undergraduate students, 30 were females and 18 were males. Their average age was 22.3 years. Participants were randomly recruited by posting recruitment advertisements on community notice boards in the University student centers, dormitories, libraries, and dining rooms, as well as sending out emails to student listserv at randomly selected departments. Each participant was compensated $15 upon completion of the experiment. They were also told in the beginning of each experiment session that the top 20% participants who saved the best sets of documents for the tasks would get an additional $10. The reason for using such an incentive mechanism was to encourage the participants to work seriously on search tasks.

3.3. Tasks

Four tasks were designed according to the faceted classification scheme developed by Li and Belkin (2008), which classifies task features on different dimensions. In the current study, four task features/facets were controlled, which were previously found to have significant influences on task difficulty in related studies. The following lists the four task facets and their definitions:

- *Product:* an intellectual task produces new ideas or findings; a factual task locates facts, data, etc.
- *Task goal (quality):* a specific task has a goal that is explicit and measurable; the goal of an amorphous task is not measurable; a combined task has both concrete and amorphous goals.
- *Complexity:* low complexity tasks involve 1–2 activities and searching in one type of information source; moderate complexity tasks involve 3–4 activities and searching in 2 types of sources; high complexity tasks involve 5 and more activities and searching at least 3 types of sources.
- *Naming:* a named task has a specified search target; an unnamed task does not have a specified search target. (Note that this feature was not in the original Li and Belkin (2008) study, but was proposed and used in Liu et al. (2011).)

The four tasks are described as below, with Table 1 listing the facet values of each task.

Task 1 (T1): Suppose you are preparing to apply for a graduate school in your major. You want to do some research online, narrow down the target schools that you think best match with your background and interest. You also want to consider their financial aid opportunities and locations when choosing the schools. After you decide on the target schools, you want to learn about their admissions requirements. **Your search task**: Find two schools that you want to apply for, and save the webpages that show their admissions requirements.

This task was an intellectual task because identifying two schools generated "findings"; its goal was combined because the two schools were amorphous, but the admissions requirements were specific; its target was unnamed, and it is a high complexity task because users would need to narrow down and identify the schools and also locate the admissions requirement pages.

Task 2 (T2): You like and admire someone so much that you want to create a Wikipedia entry for this person. **Your search task**: Please collect and save all the webpages and other online sources that are helpful for you to write such a Wikipedia entry. *Note: for the selection of the person, you want to make sure that there is not an existing entry in Wikipedia for him/her.* (The note was in the task description, and the Italic and bold font was as well.)

This task was factual because only facts (or information sources with facts) about the person were required; its goal was amorphous because the person for whom the entry was created was not specified, as were the facts/information to be included in the entry; its target was unnamed, and it is a high complexity task because users would need to try locating as many as possible sources/facts for the person.

Task 3 (T3): Your friend is a soccer fan and he is participating in a contest about soccer knowledge. For one question, he could not answer, but he had a chance to call you and ask you for information. The question is: Name at least one soccer player who had done a hat-trick in the 2010–2011 Italian Soccer League Series A. **Your search task**: Search online for the answer for the above question. Save the webpage(s) with the answer.

This task was factual because the answer was a fact; its goal was specific because there was only the right answer to look for; its target was named, and it is a low complexity task because users would just need to locate the soccer player who meets the requirements.

Task 4 (T4): You have read that Alfred North Whitehead wrote a book, Science and Modern World, and you are interested in learning more in the concept of "simple location." You would like to learn about the various interpretations and views on Whitehead's simple location as a philosophical concept. **Your search task**: Search and save 5 articles, in full-text (.pdf or .html), that you think are helpful.

This task produced an intellectual task because learning about the various interpretations and views of the concept generated "ideas"; its goal was combined because specifically 5 articles were required, but the helpfulness of these articles did not have a single measurement; its target was named, and it is a high complexity task because users would need to understand the concept and locate 5 articles that were helpful to understand the concept.

Each participant finished all 4 tasks. In order to minimum the order effect, the task order for all 32 sessions in the whole experiment followed a Latin square design.

Table 1
The four tasks and their task facet values.

Task ID	T1	T2	T3	T4
Task topic	Graduate school application	Wikipedia entry	Soccer	"Simple location" concept
Product	Intellectual	Factual	Factual	Intellectual
Goal (quality)	Combined	Amorphous	Specific	Combined
Naming	Unnamed	Unnamed	Named	Named
Complexity	High	High	Low	High

3.4. Procedure and data collection

Participants came individually to a computer lab to take part in the experiment. Upon arrival, they first completed a consent form and a background questionnaire eliciting their demographic information and search experience. After a training task, they worked with the 4 tasks in the assigned order, each for up to 15 min. Each participant was allowed to freely search online and save the documents that they thought could help them solve the assigned tasks. Users' interaction with computers was logged by Morae usability software (http://www.techsmith.com/morae.html), although the current study did not use this data, but instead relied on the questionnaire responses.

Questionnaires were administered both before and after each task to elicit users' perceived task difficulty levels, their self-assessed knowledge levels with the task topic, their interest in the topic, and so on. All ratings were based on a 7-point scale (1 = not, 4 = somewhat, 7 = extremely). They were also asked why they gave the difficulty ratings, and what made them think the task would be/was difficult. After they finished all four tasks, a final questionnaire asked if they have further comments on the experiment. They were then given the compensation.

4. Results

Difficulty ratings and reasons were collected both before and after tasks, but the current analysis considered only the post-task perceptions since they represented users' judgments based on the actual task experience. Task difficulty in this study follows the notion of Li and Belkin (2008) that it was users' subjective judgments. Users' responses to both questions about why and what made them feel the task was difficult were examined, and then classified into different categories following a coding scheme (see below).

4.1. Task difficulty rating overview

There were a total of 192 user–task sessions (48 users by 4 tasks). First examined were user ratings for the 4 tasks as an overview. A one-way ANOVA test showed that Task 4 was rated as the most difficult, which was significantly more difficult than Task 1 (Table 2). The ratings of T2 and T3 were descriptively but not statistically significantly higher than T1 and lower than T4.

It is noted that T3 is a factual, specific, named, and low complexity task (Table 1), which may have seemed to be low difficulty, but its difficulty rating was descriptively higher than T1, although not statistically significant. Without further examination, we speculate that task difficulty levels may be caused by various factors, such as users' familiarity with task topics and task type, and it may be possible that users' low topic familiarity with T3 (the most frequently given difficulty reason in T3, as shown in Table 4) made them feel that it was difficult, despite the fact that it was a low complexity task. Since the current study focuses on difficulty reasons, more examination on difficulty ratings will be conducted in future studies.

4.2. Coding scheme

Based on the task difficulty reasons given by the 32 participants at SCSU in both the pre- and the post-task questionnaires, a scheme was developed that classified different reasons into categories. An examination of the 16 participants at USC did not detect new reasons for task difficulty. In general, the various reasons spread across different aspects, including task feature, user aspect, and the interaction between the user and the task. The coding scheme is listed in Table 3. More details about this coding scheme can be found in Liu and Kim (2013).

An inter-coder reliability test was conducted by two authors using a 10% sample randomly selected in the SCSU data pool with 32 participants. The Cohen's Kappa value between the two coders was 0.87, which was satisfactory. One author coded all the 48 participants' responses to difficulty reasons.

4.3. Reasons for task difficulty perception

Table 4 lists the frequency of each difficulty reason in each task. As can be seen, some people rated the tasks as not difficult. This happened in all four tasks, with 25 cases in T1, 16 in T2, 16 in T3, and 7 in T4. The many cases for T1 being rated as not difficult correspond to the task difficulty ratings in Table 2. T1 was about locating information of the graduate schools, which could have seemed to be a familiar topic to many participants, and in fact, none of the participants listed low topic

Table 2
Mean (standard deviation) of post-task difficulty ratings in 4 tasks.

Tasks	T1	T2	T3	T4	*F(p)*
Difficulty ratings	2.75 (1.56)[*]	3.48 (1.88)	3.56 (2.01)	4.21 (1.74)[*]	**5.25 (.002)**

Bold indicates result being statistically significant.
[*] Tasks having significant differences.

Table 3
A task difficulty reason scheme.

General category		#	Category label	Examples
Not difficult		0	Not difficult	"It was not difficult." "Not hard at all." "Nothing at all."
Task feature		1	Time limitation	"Narrowing down a wide range of opportunities to just 2 schools in 15 min." "It was not really difficult except for the time taken."
		2	Complexity	"There are several components of this task that need to be addressed–this makes it more complicated."
		3	Specific requirements	"Because I'm looking for something specific, and something that isn't really popular to the rest of the world." "Because the information needed is so specific that it will be much more difficult to find."
User aspect		4	Low topic knowledge	"I have no knowledge of the subject so I won't know exactly what I'm looking for."
		5	Little experience	"Never did such a search before - or creating a Wikipedia entry for a specific person, especially someone who is not on Wikipedia already."
		6	No interest	"Topic was not all that interesting and at times, answers were difficult to find."
		7	Did not learn well in the search process	"I didn't educate myself enough about the task while searching."
User–task interaction	Uncertainty with task before starting	8	Uncertainty about task requirements	"Don't know exactly what I am supposed to be looking up."
		9	Uncertainty/Need to identify information need (decision making)	"I didn't know who to pick or where to start." "It was hard deciding on a person to pick that I was a fan of that did not already have a Wikipedia page."
		10	Uncertainty about amount of information needed	"Because I'm not sure how much information I would find on this person."
		11	Uncertainty about information available	"I am not sure if Alfred Whitehead is well known or how difficult it will be to find information related to his views."
	System selection	12	Difficulty finding the right site/system to search in	"Finding the most helpful websites."
	Query formulation	13	Hard to formulate queries	"I could not formulate the question correctly. When I would type in exactly what I was looking for other information would come up." "I did have some trouble with my search queries…"
	System performance and/or search result features	14	System performance	"I will most likely be using the school database so it may take a while."
		15	Too much (unrelated) information; too many choices	"Narrowing down a wide range of opportunities to just 2 schools in 15 min."
		16	Not enough information	"It may be very difficult to gather information on someone if they are not famous enough to have a Wikipedia page."
		17	No straightforward (exact) answer	"May not be able to get a straight forward answer."
		18	Results not as expected	"I thought just typing in the subject would just give me the straight link but it took me more time than I thought." "Going through ESPN, I figured they would have the answer, but they didn't."
	Document features	19	Need to read/comprehend information	"The amount of reading and guessing where the words "simple concept" would appear was time consuming which makes me feel the task was difficult."
		20	Resource credibility/quality	"I have to find 5 sufficient articles and I'm sure some of them will be unreliable." "This task was only difficult because I wanted to make sure I was looking at valid sources."
No specific reason		21	Does not fall in any of the above, did not state why	"Trying to find how to create a Wikipedia entry about someone is difficult."

familiarity as a difficulty reason for this task (Table 4), and most participants rated their familiarity degree of this topic as high (Table 6).

For other cases, they gave their reasons why the tasks felt difficult, and the top reasons are listed by tasks (Table 4). The cutoff frequency for each task was selected as the median value of difficulty reason frequency in that task (4, 4.5, 2, and 3.5 respectively in the four tasks). Those reasons with a frequency greater than the cutoff frequency were selected as top reasons.

Table 5 listed the top reasons for the 4 tasks. For Task 1, there were four top reasons: specific requirements, uncertainty about information need, too much (unrelated) information, and task complexity. For Task 2, the five top reasons were not enough information, specific requirements, uncertainty about information need, too much (unrelated) information, and complexity. The three top difficulty reasons for Task 3 were specific requirements, low topic knowledge, and too much (unrelated) information. For Task 4, there were 8 top difficulty reasons: low topic knowledge, too much (unrelated) information, specific requirements, need to read/comprehend information, resource credibility/quality, time limitation, not enough information, and complexity.

Table 4
Post-task difficulty reason categories and frequencies in different knowledge groups and tasks.

	Difficulty reason category	Frequency																		
		Task 1				Task 2				Task 3				Task 4				Total			
		L	M	H	T	L	M	H	T	L	M	H	T	L	M	H	T	L	M	H	T
0	*Not difficult*	4	4	17	25	6	4	6	16	10	1	5	16	6	1		7	27	10	28	65
1	Time limitation			1	1	1			1					**5**			**5**	6		1	7
2	Complexity	**2**	3	1	**6**		1	**4**	**5**					**4**			**4**	6	4	5	15
3	Specific requirements	**5**	2	6	**13**	**5**	3	**4**	**12**	**11**	1	2	**14**	**7**			**7**	28	6	12	46
4	Low topic knowledge					1			1	**15**		1	**16**	**16**			**16**	32		1	33
5	Little experience	**2**	1	1	4	1			1					3			3	6	1	1	8
6	No interest									1			1	3			3	4			4
7	Did not learn well in the search													1			1	1			1
8	Uncertainty about task requirements			1	1	1			1					1			1	2		1	3
9	Uncertainty about information need	1	1	**7**	**9**	**3**	5	1	**9**					1			1	5	6	8	19
10	Uncertainty about amount of information needed																				
11	Uncertainty about information available																				
12	Difficulty finding the right site/ system to search in													2			2	2			2
13	Hard to formulate queries									2			2	3			3	5			5
14	System performance		1	**3**	4					1			1	2			2	3	1	3	7
15	Too much (unrelated) information	1	1	**4**	**6**		2	3	**5**	**5**			**5**	**9**			**9**	15	3	7	25
16	Not enough information	1		1	2	**7**	8	**7**	**21**	1		1	2	**4**			**4**	13	8	9	30
17	No straightforward (exact) answer									1			1					1			1
18	Results not as expected		1		1						1	1	2	2			2	2	2	1	5
19	Need to read/comprehend information	1			1									**6**			**6**	7			7
20	Resource credibility/quality					2		2	4	1			1	3		1	**4**	6		3	9
21	Reason not explained			1	1			1	1											2	2
	Total	17	14	43	74	27	23	28	78	49	3	10	62	78	1	1	80	171	41	82	294

Note: 1. Blank cells mean the values were 0 s.
2. Categories 10 and 11 did not have values in the post-task difficulty reasons in all four tasks. These categories appeared in the pre-task difficulty reasons.
3. Those in bold are top reasons with high appearing frequency.

Table 5
Top reasons of task difficulty and frequency (in parenthesis) (bold for the common reason across 4 tasks).

	Post-task difficulty reasons
T1	**Specific requirements (13)** Uncertain about information need (9) **Too much (unrelated) information (6)** Complexity (6)
T2	Not enough information (21) **Specific requirements (12)** Uncertain about information need (9) **Too much (unrelated) information (5)** Complexity (5)
T3	Low topic knowledge (16) **Specific requirements (14)** **Too much (unrelated) information (5)**
T4	Low topic knowledge (16) **Too much (unrelated) information (9)** **Specific requirements (7)** Need to read/comprehend information (6) Resource credibility/quality (4) Time limitation (5) Not enough information (4) Complexity (4)

There were common reasons in the four tasks that made the users perceive the tasks as difficult: specific requirements, and too much (unrelated) information (Table 6). Other reasons leading to task difficulty in different tasks varied. T1, T2 and T4 had the same common reason of task complexity. T2 and T4 had the same common reason of not enough information. T1 and T2 had the common reason of uncertainty about information need. T3 and T4 had one common reason: low topic

Table 6
Topic knowledge groups and frequencies in the 4 tasks.

	Low knowledge	Mid-knowledge	High knowledge	Total
T1	10	8	30	48
T2	20	13	15	48
T3	36	4	8	48
T4	46	1	1	48

knowledge. T4 also had some unique reasons that were not identified as top difficulty reasons in other tasks, and they were: time limitation, need to read/comprehend information, and resource credibility/quality.

4.4. Relationships between users' background and difficulty ratings

As a frequently examined factor reflecting information searchers' background, task topic knowledge was examined regarding its correlation with task difficulty ratings. Unlike the task difficulty ratings, which better reflect users' actual experience in the post-task questionnaire, the topic knowledge in the pre-task questionnaire better reflects one's base knowledge and background with the task topic. These relationships were used in the current study.

First analyzed was the relationship between task difficulty ratings and users' topic knowledge. A significant and positive correlation was found between them ($r(128) = -.187$, $p = .034$).

In order to analyze the differences between the lower and higher topic knowledge users, they were divided into groups by their knowledge levels before they worked with the task, i.e., their pre-task topic knowledge ratings. According to the rating scale meanings, ratings 1–3 were put in a low knowledge group, 4 a mid-knowledge group, and 5–7 a high knowledge group. Table 6 shows the number of participants in each group in each task.

4.5. Task difficulty by topic knowledge group in each task

Since task topic knowledge was a factor significantly correlated with task difficulty, task difficulty reasons were also examined with different user groups that had different levels of topic knowledge. This subsection presents the main reasons why a task was difficult in different tasks for different knowledge groups of users. For each group of users, the cutoff threshold for each task was again the median frequency of difficulty reasons in that task. In the low knowledge group, the median frequency values were 1.5, 2, 1.5, and 3 respectively for the four tasks, and in the high knowledge users, those were 1, 3.5, and 1 in Tasks 1–3, while no users had high knowledge in Task 4. Difficulty reasons with a frequency greater than the cutoff values were selected as top reasons.

In these analyses, only the low and high knowledge groups were selected and compared, which would possibly reduce the chance that mid-knowledge group users could have the features of both the low and the high groups.

Table 7 shows the top difficulty reasons in the low and high knowledge groups for different tasks. As can be seen, for each task, there were some common reasons identified in the two groups, but there were also variations.

For Task 1, both low and high knowledge users identified specific requirements as a common top reason for task difficulty. Low knowledge users also identified complexity and little experience as the top reasons. High knowledge users also mentioned the uncertainty about information need, too much (unrelated) information, and system performance as top difficulty reasons.

For Task 2, two common top reasons in the two groups were not enough information and task complexity. Low knowledge users also identified the uncertainty about information need as a difficulty reason. High knowledge users mentioned task complexity as another top reason for task difficulty.

For Task 3, one common reason in both groups was specific requirements. Low knowledge users mentioned three other top reasons: low topic knowledge, too much (unrelated) information, and hard to formulate queries.

For Task 4, there was only one user who rated the knowledge level as high. One self-assessed the knowledge level as 4 (medium knowledge level), and all others ($N = 46$) rated themselves as low knowledge with the task topic. Low knowledge users mentioned many reasons for task difficulty: Low topic knowledge, too much (unrelated) information, specific requirements, need to read/comprehend information, time limitation, not enough information, and complexity. The one high knowledge user mentioned the resource credibility/quality as the task difficulty reason.

5. Discussion

5.1. Task difficulty and task type

Section 4 compared task difficulty reasons for each task. Since each task was special in topic, it would be more beneficial and valuable to compare the difficulty reasons along task types, which builds the basis, at least to some extent, for generalization of the findings. Table 8 lists the top reasons appearing in the four tasks.

Table 7
Top reasons of task difficulty and frequency (in parenthesis) in different knowledge groups (bold for common reasons in both groups).

	Low knowledge group		High knowledge group	
	N	Reasons	N	Reasons
T1	10	**Specific requirements (5)** Complexity (2) Little experience (2)	30	Uncertain about information need (7) **Specific requirements (6)** Too much (unrelated) information (4) System performance (3)
T2	20	**Not enough information (7)** **Specific requirements (5)** Uncertain about information need (3)	15	**Not enough information (7)** **Specific requirements (4)** Complexity (4)
T3	36	Low topic knowledge (15) **Specific requirements (11)** Too much (unrelated) information (5) Hard to formulate queries (2)	8	**Specific requirements (2)**
T4	46	Low topic knowledge (16) Too much (unrelated) information (9) Specific requirements (7) Need to read/comprehend information (6) Time limitation (5) Not enough information (4) Complexity (4)	1	**Resource credibility/quality (1)**

It was found that two common reasons were identified as top reasons in all tasks: specific requirements and too much (unrelated) information. The reason of "specific requirements" seems to reflect the users' wish to find the requested information right in the (top rank(s) of the) search result list, and if not, they would think that the search was difficult because it did not offer information to meet the "specific requirements". Similarly, having too much unrelated information made the users not easily find the needed information, which could have accordingly made them feel the task was difficult.

Since these two reasons appeared in all task types, they would not help in detecting task type or generalizing difficulty reasons to specific task types. What is more helpful here is to examine the top reasons that appeared in some but not all tasks, which can possibly be related with task types (features).

One of such task reasons was the uncertainty about information need. This appeared in T1 and T2. This reason of the uncertainty about information need could be due to the feature of those two tasks being unnamed in target, as well as amorphous in task goal (quality). In the graduate school application task (T1), the two schools were not named, and the task goal cannot be measured because of this unnamed nature. In the Wikipedia entry task (T2), the person for whom the entry will be created was unnamed, and the task goal was also amorphous. These features may have made the users feel uncertain about the information need, and it is indicated that tasks that are unnamed (in target) or amorphous (in task goal) would possibly make the users feel they were difficult because of the uncertainty of the information need.

Another of the common reasons appearing in some but not all tasks was task complexity, which was identified as a top difficulty reason in T1, T2 and T4. This matched the fact that these three tasks were all designed as high-complexity tasks. A task being complex in nature is likely to make the users feel it is difficult.

For both T3 and T4, users identified low topic knowledge as a top reason for task difficulty, but this was not a top reason for T1 and T2. That T3 and T4 were both "Named" in target might have allowed users to assess how much knowledge they had with the "specific" topics; on the other hand, T1 and T2 did not have their named targets, which did not provide users "specific targets" on which to assess their topic knowledge. In this case, although users can still assess their knowledge on the task topic in general, the fact that they would have to determine their specific target from this Unnamed task (i.e., the reason of "uncertain about information need", which was one of the top reasons in T1 and T2 as above mentioned) may add another layer of reason for task difficulty, which may also shadow the role of topic knowledge in leading to task difficulty.

The reason "not enough information" was identified as a top reason in T2 and T4. There was not a common task facet for these two tasks only, so it is somewhat tricky to attribute or connect this reason with the available task facets addressed in this study. A further examination of these two tasks may explain why "not enough information" was identified as a top reason. These two tasks were both high-complexity tasks, which may be a sufficient but not necessary condition to "not enough information" (e.g., T1 was also a highly complex task, but users did not have problems with finding enough information for that task). T2 asked users to collect as much information as possible for a person not having a Wikipedia page, which was a recall-based task. Not having a Wikipedia page most likely indicate that the person is not famous enough, which is also an indication of not having much information on the Internet. So it is mostly the requirements of this task, not the task facets addressed in this study, that made this task difficult. With regard to T4, it asked users to save 5 full-text articles about a specific concept of a philosopher. Despite the fact that many users were not familiar with the concept, there are a good amount of relevant articles available in library databases. However, many users relied only on general search engines, which did not return many relevant results. The underlying reason for this task being perceived difficult by users was actually not using the best search sources. The relationship between the reason of not enough information and other reasons such as not using the

right search sources can be further examined in future studies, but it can be seen that "not enough information" does not seem to connect with the addressed task facets in the current study.

5.2. Task difficulty reasons in different knowledge groups

The two groups of users with different levels of topic knowledge showed quite some differences, in addition to some commonalities, in top reasons perceived for task difficulty (Table 7). Users in both groups identified specific requirements as the common top reason in T1, specific requirements and not enough information in T2, and specific requirements in T3. As discussed above, the reason of specific requirements appeared in all four tasks, and it was also found to appear in both groups in all three tasks (Task 4 did not have high knowledge group users). In Task 2 (Wikipedia entry), the common reason of not having enough information matched the task being a recall-based one, asking users to collect all webpages and other sources (i.e., as much information as they can) for them to create a Wikipedia entry for someone.

When personalizing search results for users, it would be more interesting and beneficial to look at the differences in difficulty reasons between the two groups. This discussion is divided by tasks and continued below.

For Task 1 (graduate school application), low knowledge users identified two top reasons that high knowledge users did not have: complexity and little experience. On the other hand, high knowledge users identified three top reasons that low knowledge users did not have: the uncertainty about information need, too much (unrelated) information, and system performance. For the low knowledge users, task complexity and little experience with the task were both general perceptions with the task as a whole, as compared with the reasons identified by the high knowledge users, uncertain about information need, too much (unrelated) information and system performance, which are all more specific and address the target schools in mind as well as the search results. This may be because the high knowledge users knew better what to do with the task than the low knowledge users, who may not have clear ideas how to deal with such tasks.

For the high knowledge users, the identified top reason of uncertainty about information need seems to indicate that, in this case, the more the user knows about a topic, the more likely that he/she feels uncertain about the information need. Those who know little about the topic may not even realize the information needs they could have had. Considering the task facets aspect, this task was intellectual, with combined goal (quality) (both amorphous and specific aspects appeared in the task), the target was unnamed, and it was a highly complex task. The amorphous goal and unnamed target of this task could have made those who had some topic knowledge feel uncertain about the information need. As for the reasons of "too much (unrelated) information" and "system performance", similar discussions can be made. That the users have a certain level of knowledge may allow them to judge the results as not relevant. On the other hand, the low knowledge users may not have advanced this far yet, and so they were only feeling the task complexity in a general sense and did not have enough experience in dealing with this kind of task. Further examination and confirmation on these speculations could be performed in future studies.

One implication of the differences in findings from user groups in T1 is that for low knowledge users, when they work with the tasks that have amorphous goals, an unnamed target, and call for an intellectual product, the system can try to make them feel experienced, or have some background with this topic, for example, by suggesting what others usually look for and collect information. The system can also try to decompose the tasks for the users, making them feel the task is less complex, for example, by organizing the information for users while they search. For high knowledge users, when they work with these types of tasks, the system can help them to identify or narrow down their information need. There could be various approaches to this goal such as offering options in categories, and providing filtering functions to narrowing down. Meanwhile, it is acknowledged that the system will need to somehow know the task type beforehand. This is an issue that has attracted research attention, for example, Liu , Belkin, et al. (2012) tried to predict task type by users' behaviors.

For Task 2, low knowledge users identified uncertainty about the information need (to identify the person for whom the entry will be created) as a top reason, but high knowledge users did not. On the other hand, high knowledge users identified task complexity as a top reason, but the low knowledge users did not. The pattern was seemingly opposite than that in T1, in which low knowledge users felt the task was complex but high knowledge users felt uncertain about the information need. However, a closer look at the two tasks and the task difficulty reasons indicates that the patterns are not opposite for no

Table 8
Common difficulty reasons by tasks.

	T1	T2	T3	T4	Task facet
Specific requirements	x	x	x	x	n/a
Too much (unrelated) information	x	x	x	x	n/a
Uncertain about information need	x	x			Unnamed/amorphous goal
Complexity	x	x		x	Complexity
Low topic knowledge			x	x	Named
Not enough information		x		x	**
Need to read/comprehend information				x	n/a
Resource credibility/quality				x	n/a
Time limitation				x	n/a

n/a: Task difficulty reasons appeared in all four tasks, or only one task, making it not applicable to discuss task facet.
** : The task difficulty reason appeared in two tasks with no common facet.

reason. Different than T1 which was an intellectual task and asked users to collect two schools' admissions pages (specific and amorphous), T2 was factual and asked users to collect all the information needed to create the Wikipedia entry (amorphous). Looking at the two tasks, what made them complex was a bit different: for T1, users had to identify the schools and also collect information, for T2, users had to identify the person, collect information, and more importantly, collect as many sources as possible. In this task, the low knowledge users perceived the task difficulty because they had to identify the person (uncertain about information need), but high knowledge users showed a further understanding of the task requirement by giving task complexity as a top difficulty reason (collect as much information as possible). In summary of both T1 and T2, high knowledge users seemed to be able to identify the more specific reasons why a task was difficult than low knowledge users. This also seems to require considering the task features/types when discussing the task difficulty reasons, which confirmed the importance of task type in task difficulty research.

The implication for this on system design is that for amorphous and unnamed factual tasks that are recall-based (different than the intellectual tasks such as Task 1), it would be helpful to have features/functions to help users save the sources, in addition to providing as many related sources about the topic as possible. This would be done in hopes to reduce users' perception of task difficulty caused by task complexity, especially for those high knowledge users.

For Task 3, high knowledge users did not have other top reasons for task difficulty besides the common reason of "specific requirements". Low knowledge users identified three other top reasons: low topic knowledge, too much (unrelated) information, and hard to formulate queries. The low knowledge reason matches their low knowledge background. For low knowledge users, it seems reasonable that they may feel difficulty in formulating queries, and accordingly, the results may not be ideal or relevant, leading them to feel there were "too much unrelated information". Based on the findings in T3, the implication for system design could be to provide easily read results (or results matching with users' knowledge gain in the task accomplishment process), in addition to query suggestions, especially to the low knowledge users.

5.3. Contributions, limitations, and future directions

This paper reports an examination of task difficulty reasons as seen in different task types and user knowledge groups. The research confirmed a task difficulty reason scheme developed by Liu and Kim (2013). Furthermore, the original merits of this paper are the analyses of task difficulty reason in different task types and user groups. These contribute to personalize task difficulty reasons by task and user groups.

One limitation of the current research is the nature of the controlled lab experiment. The findings are important in that they explored the relationships between task difficulty reasons and task type. Meanwhile, it is acknowledged that task difficulty reasons may have been limited, to some extent, to the controlled lab experiment and the assigned task nature in this experiment, although we think, on the other hand, that the design of the various task types following the task facet scheme could have taken care of most frequently seen task difficulty reasons. Future studies will collect task difficulty reasons in naturalistic settings.

Besides the above future plan and the implications that have been discussed about the task difficulty scheme, which can be linked with future research, another future direction is to examine if there are relationships between task difficulty reasons, users' search behaviors, and contextual factors including tasks and knowledge background. This can provide us with a better understanding of user behaviors, as well as ways to predict task difficulty reasons based on user behaviors.

6. Conclusions

Based on the data collected in a controlled lab experiment, the paper explored why information searchers felt an information task was difficult and identified the various reasons that led to task difficulty perception. A task difficulty classification scheme previously developed by Liu and Kim (2013) was confirmed. It involved several aspects: users, tasks, and user–task interaction. Following this scheme, the difficulty reasons that users gave for each task session were categorized. Difficulty reasons were compared in different tasks and task types. It was found that the common reasons leading to task difficulty in all four tasks were specific requirements and too much (unrelated) information, meaning that users wanted the system to provide the right, exact, and relevant information in the search result pages. Various difficulty reasons could be linked with the different task types. The study also examined the commonalities and differences between the identified top difficulty reasons between users with different levels of topic knowledge. The possible reasons that led to the different top reasons were discussed, and some implications were provided for changes in system design to personalize searches for different groups of users, besides helping reduce task difficulty for all users in general. Future research can examine the relationships between task difficulty reasons, users' search behaviors, and contextual factors (tasks, knowledge background, etc.). Future studies can also attempt to collect data and identify the task difficulty reasons in a naturalistic setting, which all help improve our understanding of task difficulty and building IR systems that help reduce users' perception of difficulty.

References

Allen, B. L. (1991). Topic knowledge and online catalog search formulation. *Library Quarterly, 61*(2), 188–213.

Arguello, J. (2014). Predicting search task difficulty. In *Proceedings of the European Conference on Information Retrieval (ECIR) 2014*.

Aula, A., Khan, R., & Guan, Z. (2010). How does search behavior change as search becomes more difficult? *Proceedings of the International Conference on Human Factors in Computing Systems (CHI), 2010*, 35–44.

Byström, K. (2002). Information and information sources in tasks of varying complexity. *Journal of the American Society for Information Science and Technology, 53*(7), 581–591.

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

Carmel, D., Yom-Tov, E., Darlow, A., & Pelleg, D. (2006). What makes a query difficult? In *Proceedings of the 27th international conference on research and development in information retrieval (SIGIR) 2006*.

Cronen-Townsend, S., Zhou, Y., & Croft, W. B. (2002). Predicting query performance. In *Proceedings of the 23th international conference on research and development in information retrieval (SIGIR) 2002*.

Duggan, G. B., & Payne, S. J. (2008). Knowledge in the head and on the web: Using topic expertise to aid search. *Proceedings of the International Conference on Human Factors in Computing Systems (CHI), 2008*, 39–48.

Gwizdka, J. (2008). Task difficulty revisited. In *Proceedings of the annual conference of the American Society for Information Science & Technology (ASIS&T) 2008*.

Gwizdka, J., & Spence, I. (2006). What can searching behavior tell us about the difficulty of information tasks? A study of Web navigation. In *Proceedings of the annual conference of the American Society for Information Science & Technology (ASIS&T) 2006*.

Hembrooke, H. A., Granka, L. A., Gay, G. K., & Liddy, E. D. (2005). The effects of expertise and feedback on search term selection and subsequent learning. *Journal of the American Society for Information Science & Technology, 56*(8), 861–871.

Hsieh-Yee, I. (1993). Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the American Society for Information Science, 44*(3), 161–174.

Kellar, M., Watters, C., & Shepherd, M. (2007). A field study characterizing Web-based information-seeking tasks. *Journal of the American Society for Information Science and Technology, 58*(7), 999–1018.

Kelly, D., & Cool, C. (2002). The effects of topic familiarity on information search behavior. In *Proceedings of the 10th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL) 2002* (pp. 74–75).

Kim, J. (2006). Task difficulty as a predictor and indicator of web searching interaction. *CHI '06*, 959–964.

Li, Y. (2008). Relationships among work tasks, search tasks, and interactive information searching behavior. Unpublished dissertation. Rutgers University.

Li, Y., & Belkin, N. J. (2008). A faceted approach to conceptualizing tasks in information seeking. *Information Processing & Management, 44*, 1822–1837.

Liu, J., & Belkin, N. J. (2010). Personalizing information retrieval for multi-session tasks: The roles of task stage and task type. In *Proceedings of the 31th international conference on research and development in information retrieval (SIGIR)* (pp. 26–33).

Liu, J., & Kim, C. (2013). Why do users perceive search task as difficult? Exploring difficulty in different task types. In *Proceedings of the symposium on Human Computer Interaction and Information Retrieval (HCIR) 2013*.

Liu, J., Cole, M., Liu, C., Bierig, R., Gwizdka, J., Belkin, N. J., et al. (2010). Search behaviors in different task types. In *Proceedings of the 10th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL) 2010*.

Liu, J., Gwizdka, J., Liu, C., & Belkin, N. J. (2010). Predicting task difficulty for different task types. In *Proceedings of the annual conference of the American Society for Information Science & Technology (ASIS&T) 2010*.

Liu, J., Liu, C., Yuan, X., Belkin, N. J. (2011). Understanding searchers' perception of task difficulty: Relationships with task difficulty. In *Proceedings of the annual conference of the American Society for Information Science & Technology (ASIS&T) 2011*.

Liu, C., Belkin, N. J., & Cole, M. (2012). Personalization of search results using interaction behaviors in search sessions. In *Proceedings of the 33rd international conference on research and development in information retrieval (SIGIR) 2012*.

Liu, J., Liu, C., Cole, M., Belkin, N. J., & Zhang, X. (2012). Examining and predicting search task difficulty. In *Proceedings of the 14th international Conference on Information and Knowledge Management (CIKM) 2012*.

Marchionini, G. (1989). Information-seeking strategies of novices using a full-text electronic encyclopedia. *Journal of the American Society for Information Science, 40*(1), 54–66.

Qiu, L. (1993). Analytical searching vs. browsing in hypertext information retrieval systems. *Canadian Journal of Information and Library Science, 18*(4), 1–13.

Sihvonen, A., & Vakkari, P. (2004). Subject knowledge improves interactive query expansion assisted by a thesaurus. *Journal of Documentation, 60*(6), 673–690.

Toms, E., MacKenzie, T., Jordan, C., O'Brien, H., Freund, L., Toze, S., et al. (2007). How task affects information search. In N. Fuhr, N. Lalmas, & A. Trotman (Eds.), *Workshop pre-proceedings in initiative for the evaluation of XML retrieval (INEX) 2007* (pp. 337–341).

Vakkari, P. (1999). Task complexity, problem structure and information actions: Integrating studies on information seeking and retrieval. *Information Processing & Management, 35*(6), 819–837.

Vakkari, P., Pennanen, M., & Serola, S. (2003). Changes of search terms and tactics while writing a research proposal: A longitudinal research. *Information Processing & Management, 39*(3), 445–463.

White, R. W., Dumais, S., & Teevan, J. (2009). Characterizing the influence of domain expertise on Web search behavior. In *Proceedings of the 2nd international conference on Web Search and Data Mining (WSDM) 2009* (pp. 132–141).

Wildemuth, B. (2004). The effects of domain knowledge on search tactic formulation. *Journal of the American Society for Information Science and Technology, 55*(3), 246–258.

Zhang, X., Anghelescu, H. G. B., & Yuan, X. (2005). Domain knowledge, search behavior, and search effectiveness of engineering and science students. *Information Research, 10*(2), 21.

Zhang, X., Liu, J., & Cole, M. J. (2013). Task topic knowledge vs. background domain knowledge: Impact of two types of knowledge on user search performance. *Advances in Intelligent Systems and Computing, 206*, 179–191.