Information Processing and Management 57 (2020) 102069

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Implicit information need as explicit problems, help, and behavioral signals

Shawon Sarkar*, Matthew Mitsui, Jiqun Liu, Chirag Shah

School of Communication and Information, Rutgers, The State University of New Jersey, New Brunswick, NJ 08901, USA

ARTICLE INFO

Keywords:
Information need
Information searching
Interactive IR

ABSTRACT

Information need is one of the most fundamental aspects of information seeking, which traditionally conceptualizes as the initiation phase of an individual's information seeking behavior. However, the very elusive and inexpressible nature of information need makes it hard to elicit from the information seeker or to extract through an automated process. One approach to understanding how a person realizes and expresses information need is to observe their seeking behaviors, to engage processes with information retrieval systems, and to focus on situated performative actions. Using Dervin's Sense-Making theory and conceptualization of information need based on existing studies, the work reported here tries to understand and explore the concept of information need from a fresh methodological perspective by examining users' perceived barriers and desired helps in different stages of information search episodes through the analyses of various implicit and explicit user search behaviors. In a controlled lab study, each participant performed three simulated online information search tasks. Participants' implicit behaviors were collected through search logs, and explicit feedback was elicited through pre-task and post-task questionnaires. A total of 208 query segments were logged, along with users' annotations on perceived problems and help. Data collected from the study was analyzed by applying both quantitative and qualitative methods. The findings identified several behaviors – such as the number of bookmarks, query length, number of the unique queries, time spent on search results observed in the previous segment, the current segment, and throughout the session – strongly associated with participants' perceived barriers and help needed. The findings also showed that it is possible to build accurate predictive models to infer perceived problems of articulation of queries, useless and irrelevant information, and unavailability of information from users' previous segment, current segment, and whole session behaviors. The findings also demonstrated that by combining perceived problem(s) and search behavioral features, it was possible to infer users' needed help(s) in search with a certain level of accuracy (78%).

1. Introduction

Information need is one of the most extensively studied constructs of information seeking behavior (Case, 2012). Existing studies have explored the concept of information need from different perspectives (e.g., Belkin, 1980; Belkin, Oddy, & Brooks, 1982a; 1982b; Cole, 2011; Dervin & Nilan, 1986; Kuhlthau, 1991a; Savolainen, 2012; Taylor, 1962; 1968; Wilson, 1981), various aspects related to it (e.g., Afzal, 2017; Al-Fedaghi, 2008; Hoenkamp, 2015; Julien & Duggan, 2000; Ruthven, 2019a), information need of individuals or

* Corresponding author.
E-mail address: shawon.sarkar@rutgers.edu (S. Sarkar).

https://doi.org/10.1016/j.ipm.2019.102069
Received 16 January 2019; Received in revised form 21 June 2019; Accepted 27 June 2019
Available online 03 July 2019
0306-4573/

a particular population (e.g., Dutta, 2009; Ikoja-Odongo & Ocholla, 2003), and different theoretical information seeking models related to need such as the reference interview model by Taylor (1968), the ASK model of Belkin et al. (1982a,b), sense-making approach of Dervin and Nilan (1986) and Dervin (1992), Bates' berry-picking model (2002) and so on. However, the very elusive and sometimes verbally inexpressible nature of information need makes it harder to grasp and define (Cole, 2011; Wilson, 1981). Often, what is required to fulfill an information need is not clear to the researchers as well as the individual seeker (Cole, 2011).

Thus, based on existing theoretical models of information behavior, the study we report here has adopted a holistic approach to understand and explore the concept of information need from a fresh methodological perspective. Towards this objective, we examined users' perceived barriers and desired help at different stages of information search by analyzing various implicit and explicit user search behaviors. This, in turn, allowed us to focus on representing a user's unexpressed information need in terms of those explicit behavioral signals. We collected explicit feedback from information seekers about their perceived barriers or problems and the help that they sought at that given moment, with an assumption that their information need is strongly connected to those two aspects of information seeking (Bates, 1989; Cole, 2011; Dervin, 1983b; Dervin & Nilan, 1986; Kari, 1998). With that in mind, we planned to investigate the following research questions (RQs) that link the study of behavioral data/signals and problems and help.

1. To what extent can the problems or barriers perceived by an individual in different stages of information search be predicted from different search behaviors?
2. To what extent can the best help for an individual be predicted from the search behaviors and the perceived barriers/problems at a given moment in the information search process?

To address these research questions, the work reported here explores various characteristics of tasks, information seekers' implicit behaviors, and explicit feedback in the course of online information search episodes including the problem(s) they face and suggestion for help(s) to address solutions to the potential difficulties in finding information. We focused on the information searching context which is the micro-level of seeking behavior where users interact with IR systems to achieve their goals.

1.1. Significance and implication of the current study

Information need that arises within a person's mind is the main initiator of information seeking behaviors (Case, 2012). This information need can be innate or external based on the person's situation (e.g., cognitive, affective, or social), and that situation is the starting point of the information seeking process (Dervin & Nilan, 1986). Therefore, one may argue that an information need cannot be measured or concretely observed. In other words, it cannot be quantified because it is a behavioral and cognitive state of a person, and there is no empirical way to determine if what one is experiencing is a need to know something. However, the current study argues that information need is a cognitive state that is expressly felt and immediately recognizable, and therefore can be observed and measured by observing other expressible physical and cognitive activities such as the problems they face and helps they need to make sense of the problem while seeking information. Thus, one of the key contributions of this work is a new methodology to understand and infer information need by examining the related aspects, such as "problems" and "helps". This work is an experimental study about a new approach, and this article reports the findings, strengths and weaknesses identified in this approach. Use of such methodological approaches based on problem-help attributes has not been previously explored for investigating information need and thus should provide value to conceptualize and operationalize the construct of information need, especially in the information search context. Furthermore, there has been a move in interactive information retrieval research to examine how task and user context affect their commonly observable seeking behaviors and strategies (e.g., Jiang, He, & Allan, 2014; Liu et al., 2010), however, while the perceived problems and helps are a product of these contexts and strategies, the nature of these relationships has mostly been unexplored. Thus, these aspects distinguish our current study from earlier studies which applied various methodological and theoretical approaches to explore the concept of information need. In addition, our approach is also interpretive and focuses on situated sense-making processes during information search episodes to extract a way of identifying information need behind them.

Taking this further, the work reported here shows how to explicate potential current and future problems and preferred help within information search episodes using behavioral data. This knowledge can then be used to provide more personalized recommendations to an information seeker, as well as to preempt potential problems by offering help before the searcher even realizes the need for such help. The rest of this article is organized as follows. Section 2 discusses the research on information need and concepts of problem, help, and other constructs that have been used in our study, such as tasks and behavioral measures to establish the background and introduces a conceptual framework for the current study. Section 3 discusses how we incorporated these constructs into the study design. Sections 4 and 5 describe the methodology, data analysis and the results followed by the discussion of the results and concluding remarks in the subsequent sections.

2. Background and theoretical frameworks

We begin with a discussion of information need in the online information search context. An information need is an evolving state of mind at a given moment in a sequence of time-space moments in a person's life within which the person engages in seeking information through interconnected cognitive and behavioral actions. Although it is an abstract state of mind, it is felt and recognizable, and therefore can be observed and measured by observing other expressible physical and cognitive activities such as the problems users face and helps they need to make sense of the problem while seeking information. This section discusses some of the existing research on information need and related aspects that inspired this study.

2.1. The concept of information need

In the 1960s, Taylor (1962, 1968) presented a cognitive model of question-negotiations where he described that the human process of question formulation goes through four stages of information need: the actual but unexpressed need for information; the conscious, within-brain description of the need; the formalized need expressible in words; and the compromised need which is the question formulated in spoken language for the information system (Taylor, 1968). Thus, Taylor's four levels of information need can be divided into two parts: a user's original innate information need which is a psychological state of mind not expressed verbally and a user's process of formulate his/her need in a question for an intermediary or an information system (Lundh, 2010). The compromised need may be a question asked to a librarian or a query statement entered into an IR system. Belkin et al. (1982a,b) furthered Taylor's theory of visceral and conscious stages of information need by proposing the concept of "Anomalous State of Knowledge" (ASK), where they described information need as an uncertain and unspecified problem or a gap between the user's knowledge about the problematic situation and what information the user needs to know to solve the problem. Similarly, Dervin (1976, 1983a, 1992) visualized information need as a gap in a user's knowledge in a situation when the user must create a "new sense" of the situation by filling the gap with information or help to move forward in the information seeking process.

Recent years have witnessed growing research efforts on characterizing and supporting information needs in search interactions (e.g. Moshfeghi, Triantafillou, & Pollick, 2016; Ruthven, 2019b; White, Jose, & Ruthven, 2003a). Based on Taylor's four levels of need (Taylor, 1968) and Belkin's anomalous state of knowledge (Belkin, Oddy, & Brooks, 1982b), many studies conceptualize information need in information retrieval context as a well-defined question which articulates users' requirements into queries to the IR system (e.g., Cole, 2011; Hoenkamp, 2009; 2015; Jansen, 2010). Hoenkamp (2009) proposed a formal definition of the ubiquitous notion of information need in the context of IR where the pair of query and selected relevant document seems to act as dual descriptions of an information need. On the other hand, other studies such as Savolainen (2012) emphasized the contexts surrounding an information need as a "jointly constructed understanding about the extent to which additional information is required to make sense of the issue at hand" (Savolainen, 2012). Broder (2002) created a taxonomy of Web search goals, and identified three types of "need behind the query": *Navigational* - the immediate intent is to reach a particular site; *Informational* - the intent is to acquire some information assumed to be present on one or more Web pages; *Transactional* - the intent is to perform some web-mediated activity. Cole, Leide, Beheshti, Large, and Brooks (2005) discussed known item and unknown item information needs. A known information need is a useful state to start seeking information because users know what information is needed. In the case of unknown information needs, users only have a vague conception of their needs and what information is needed (Cole et al., 2005). Thus, in their own ways, existing studies have called for attention to the concept and representation of information need. Therefore, information needs can be recognized and measured by observing multiple aspects of the search process, such as the different queries that are formulated or the different retrieved documents user subsequently select as relevant.

2.2. Problems encountered during search, and desired helps

Previous studies have shown that individuals frequently face difficulties when looking for information (e.g., Feild, Allan, & Jones, 2010; Savolainen, 2015; 2016) which cause uncertainties during an information seeking process (Chowdhury, Gibb, & Landoni, 2011). These include internal barriers (e.g., lack of knowledge, unable to articulate and express the need) (Belkin et al., 1982b), external barriers (e.g., time constraints, institutional restrictions) (Savolainen, 1995), and interpersonal barriers (e.g., lack of help from other people) (Swigon, 2011), and other barriers as well (Chowdhury et al., 2011). In their study, Chowdhury et al. (2011) identified twenty-one problems or barriers that people might face during information seeking processes such as information overload, irrelevant search results, and so on. When people encounter problems in the search process, they desire help or assistance either from an IR system or a human to solve problems (Jansen, 2006). Xie and Cool (2009) investigated help-seeking behaviors or problematic situations that arise in searching digital libraries, and identified fifteen types of help-seeking problematic situations that lead novice users to look for help. They also identified possible causes behind those situations such as users' knowledge structure related to the domain of the search tasks, system knowledge, and information-retrieval knowledge. They showed that help-seeking situations are heavily influenced by users' personal and cognitive characteristics as well as the nature of task (Xie & Cool, 2009). These help-seeking situations and barriers that users face provided an appropriate framework for the current study.

2.2.1. Explicit and implicit behavioral measures

In information seeking and interactive IR research communities, existing studies have used various search behavior data as indicators of users' interests, intentions, and level of satisfaction in the investigations of search interactions (e.g., Fox, Karnawat, Mydland, Dumais, & White, 2005; Liu et al., 2015; Mao et al., 2016; Pirolli & Card, 1999; Rha, Mitsui, Belkin, & Shah, 2016; Shah & González-Ibáñez, 2011; White & Huang, 2010; White, Jose, & Ruthven, 2006), such as query reformulation behavior, the number of web pages visited in a search session, the number of unique web pages visited, the number of times users clicked on the browser's back button in each task, the time spent by a searcher on each web page (dwell time), the total time on each task (time), and many more (Liu et al., 2010). Shah and González-Ibáñez (2011) measured the relevance of a web page based on participants' bookmarks or from where one or more snippets were collected. The study of query diversity, i.e., how many different queries that a user generates during a search session to find information on the same topic, is also an important measure to explore user need satisfaction, as well as how they perceive and approach the task they need to fulfill (Rha et al., 2016; Shah & González-Ibáñez, 2011). Rha et al. (2016) proposed seven types of query reformulation – *Generalization (1), Specialization (2), Word Substitution (3), Repeat (4), New (5), Spelling Correction (6), and Stem Identical (7)*. In a different study, Saracevic and Kantor (1988) examined the number of query terms

generated by users to assess the complexity of users' task. From a different approach, some studies have used implicit behavioral measures such as mouse clicks, keyboard actions, scrolling count, number of websites visited, dwell time, browsing patterns, image count, page size, queries and search engine result pages (SERPs) to detect changes in users' information needs (e.g., Fox et al., 2005; White, Jose, & Ruthven, 2003b; 2006). These studies have found that a combination of these implicit measures could provide good predictions of explicit judgments of user satisfaction. Therefore, these subtle user activities within the information search process can give hints from which it is possible to construct a general pattern of users' decision-making behavior as well as their information needs. The objective of this study is to investigate whether it is possible to predict problems that a user is encountering and helps he or she prefers from various user activities in a search session, such as time spent on a web page, the order of visit to a web page in a search session, frequency and number of query changes, and many more.

2.3. User characteristics

A searcher's state of knowledge is a major factor in the information search process that determines interactions with IR systems in retrieving relevant and useful information. According to Marchionini (1997), "personal information infrastructures" consist of users' general cognitive abilities and their knowledge of the task domain and of searching skills for IR systems. Existing research has identified that a user's knowledge structure influences search strategies and determines success or failure of information retrieval. For example, users' knowledge about the task domain or problem shapes their understanding of the problem as well as their expectations about relevant information and possible solutions, hence influencing search behaviors and performance (Dumais & Belkin, 2005; Marchionini, Dwiggins, Katz, & Lin, 1993). Furthermore, users' perception of the difficulty of a search task influences the way they pursue the task (Bell & Ruthven, 2004). If they perceive a task as an easy one or a complex one, they may interact with the IR system or pay attention less or more accordingly. Knowledge and experience about the IR system and search process assist users to develop conceptual and procedural strategies. Aligned to these findings, we considered users' knowledge structure as important aspects in our study design.

2.4. Task characteristics

Kelly (2009) argued that most information needs can be characterized in terms of task and topic; she defined a task as a representation of the goal or purpose of the search, and a topic as a representation of the subject area that is the focus of the task. Thus, specific tasks not only lead to information searching, but dimensions of tasks also influence how users search for information. The nature and types of tasks, time and stages affect how users plan and apply different types of search strategies in the search process (Xie & Cool, 2009). Amongst task characteristics, the complexity of the task (low, moderate or high) has been recognized to be an important factor that affects users' information-seeking strategies (Byström, 2002; Vakkari, 1999). Various studies conceptualized task complexity from various points of view (Bell & Ruthven, 2004; Byström, 2002; Saracevic & Kantor, 1988; Vakkari, 1999). Byström and Järvelin (1995) construed task complexity as prior knowledge of a task and expected outcomes. According to them, as task complexity or difficulty increases, users' needs for sources of information, domain information, and problem-solving information also increase. Marchionini (2010) found that users spent more time and performed more moves for an open-ended task (there is no one correct answer) than for a closed task (with only one correct answer). Capra, Arguello, O'Brien, Li, and Choi (2018) examined task complexity through the lens of a priori determinability and found that specifying items and dimensions of tasks often generate different effects on task determinability (uncertainty). Li and Belkin (2008) examined the effect of intellectual vs. decision/solution tasks on user search behaviors. Intellectual tasks were found to involve more result pages viewed, longer queries, and higher self-ratings on task success. Therefore, all these aspects of the task affect the task performance process, and all are crucial for the analysis and evaluation of the relationship between information need and information seeking behavior.

Although it is difficult to model users' information need for an experimental study, it is possible to simulate an information need with the combination of the specific task and topic (Kelly, 2009) by using a *simulated work task* (Borlund, 2003; 2016; Borlund & Ingwersen, 1997) where researchers provide a pre-designed task description to participants in laboratory studies. Each task is designed to elicit a "simulated information need" (Borlund, 2003; 2016). Along these lines, we situated tasks in the central position in relation to the problem of unspecified need, aiming to investigate the relationship between tasks and users' information searching behavior. Our ultimate goal is to advance the understanding of barriers that people face in online information seeking at the task level and the helps they would prefer and eventually improve the existing search systems or create a new one for successful information searching experiences. Within Dervin's Sense-Making and Gap-bridging framework, the problems and gaps perceived at a given moment may give rise to a searcher's perception of what resources would help at any given moment. Querying and browsing behaviors in Web search can be seen as gap-bridging strategies.

3. Implementation of the frameworks in the experimental design

Based on the socio-cognitive construct of information need established by above-mentioned studies, we envision a user's information need as the perceived inadequacies or a gap in the user's conceptual state of knowledge, which changes continuously as the user progresses through the information seeking process. When users face a problematic situation and do not have adequate knowledge to solve it, they may have an ill-defined information need which is not clear enough for articulation. They need additional information to clarify their thoughts. As the users move through the information seeking process, their need evolves, and gradually they have clearer idea about what information they need. They must also compromise the need, and their judgments of information

also change, reflecting the change in their knowledge and understanding of the problem (Kuhlthau & Vakkari, 1999; Kuhlthau, 1991b; Saracevic, 1975; Xie & Cool, 2009). Aligned with these studies, our approach also considers information need as an elusive and intuitive construct and thus unknowable and hard to articulate in a query to an information system (Belkin et al., 1982b). A realization of need in one's mind may differ from its formal expression in words. Moreover, as problematic situations arise in social settings and information is sought and used in social situations; we assume that the need for information also arises from an individual's sociological environment. Since humans' cognitive states and socio-cultural settings and situations vary; differences in these characteristics can give rise to different information needs. Therefore, the need has to be inferred from the visible actions made by the seeker during an information seeking process (Belkin & Vickery, 1985; Case, 2012), and seeker's perceptions and interpretations of their own information needs, and how they choose, formulate, and express their needs. To do that, we have drawn on the above-mentioned theoretical and methodological works and attributes of Dervin's sense-making approach (i.e., situation-gap-use) (Dervin, 1992; Dervin, Foreman-Wernet, & Lauterbach, 2003; Kari, 1998; Reinhard & Dervin, 2012) to map information need in different stages of an information search process. The mapping specifies various kinds of situations that create needs for information or "gaps" (Kari, 1998) based on problems or barriers that the users perceive when expressing those needs and while seeking information, helps that the users prefer to solve the problem and other actions that users perform such as the queries that are formulated, or the retrieved documents user subsequently select. When a user faces a gap during the information search process, he or she begins to ponder on possible problems, and possible gap-bridging strategies (preferred helps). These judgments are probably based on his/her past experiences, beliefs, and assumptions anticipated with them. The core of our study is the concept of "gap" or the "assumptions of discontinuity," and individuals bridging the gaps while moving through a given time and space by continuously seeking and making sense of the real world (Dervin et al., 2003). This continuous communication between humans and the real world can also be observable in the communication between individuals and IR systems on a micro level while individuals engage in information search processes (Cole, 2011). Thus, our approach conceptualizes the following main factors involved in constituting how people realize and express their information need:

- *Situation* - a given moment in an information search episode in which a user needs to make sense of a gap. In a broader sense, this also provides the context for the information searching process such as user's cognitive and socio-demographic features as well as the characteristics of tasks
- *Gap or need* - represents the information need to clarify a problematic situation
- *Barriers* - factors which the individual perceives as problems or constraints in the information seeking process
- *Seeking behaviors* - consisting of users' implicit and explicit behavioral actions such as seeking strategies, query formulations, source selections and so on.

Fig. 1 explains the problem-help model of information need and explains relationships among these concepts.

In a controlled experimental study, regardless of which type is used, the nature of the task and the manner in which they are described and presented to the participants may affect the results of the study (Wildemuth, Freund, & Toms, 2014). Various task characteristics affect how users seek and use information (Talja & Nyce, 2015) and thus, in turn, affect the experimental study. Therefore, it is essential to measure and record influences of tasks on users' information seeking behavior. Our study tries to achieve

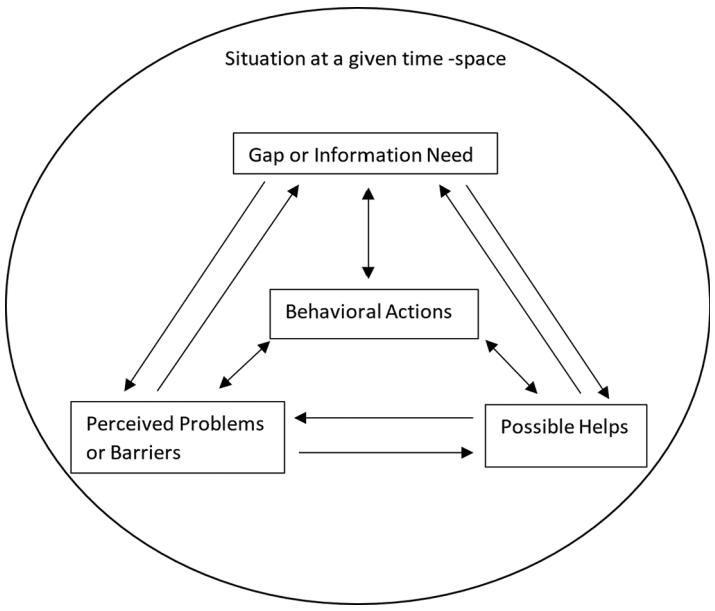

Fig. 1. An expanded model of sense-making in information seeking. Adopted from Kari (1998) and Dervin et al. (2003).

that by taking explicit user feedback (e.g., questionnaire, survey), or implicit, contextual data such as dwell time, topic knowledge, and task information (Liu & Belkin, 2010). Based on these assumptions, we designed the experimental framework for the study to observe various aspects of the search process, and discuss the findings in light of them.

4. Methodology

To address the research gaps and questions discussed above, we designed a controlled laboratory-based user study in a research university environment using a mixed method approach, where participants were asked to perform three search tasks. Participants were additionally asked to report problems faced at various steps in the search process, as well as what would help them at that moment in the process. The main components of the study procedure are detailed in the following subsections.

4.1. Participants

Seventeen participants were recruited from the undergraduate student population at a United States research university via various channels, including advertisements to emails lists, Facebook groups and posting announcements around the campus. Participants registered online through the study registration website. Each participant was compensated $20 in cash after completing the entire study. Participants came from diverse educational backgrounds ranging from History, Arts, to Computer Science, Psychology, Public Health and so on. Their age ranged from 18 to 22 years, with an average age of 20.2. 64.7% were females, and all of them were native English speakers. Most participants reported that they have more than ten years of web search experience.

4.2. Study procedure

Participants conducted their web searching using Google Chrome in a desktop computer equipped with necessary applications. Participants began by reading a description of the study procedure, browser interface features, and Chrome extension customized for them. Participants then performed their three search tasks, where they were required to search for information on the Web, based on the given simulated task scenarios. Each time they opened a new tab to begin formulating/reformulating a query, the Chrome extension provided a problem/help questionnaire (descriptions are provided in the *Information seeking barriers and help* section below). Before and after each search task, participants answered short questionnaires on task familiarity, experience, perceived difficulty, and topic familiarity. The study sessions – including participants' interactions such as web pages visits, bookmarks, queries, and problem/help answers - were captured through a custom Chrome extension and Morae.[1] After completing all tasks, participants' sessions concluded with a brief semi-structured exit interview about their levels of satisfaction with their search sessions, their evaluation of the search process, problems they encountered, desired helps, perceptions of their search experiences in everyday lives, and their expectations of an ideal information retrieval system performance. The length of a laboratory session for each participant ranged from 50 to 80 min total. Fig. 2 depicts the structure of pre-task, search session, and post-task for a single task.

4.3. Study design and data collection

This section discusses the main components of the study design.

4.3.1. User characteristics and states of knowledge

As users' own states of knowledge is a major factor in their search processes, we collected demographic characteristics such as age, gender, educational background, English proficiency, and search experience on the web environment at the time of the registration (we report the data in the Participants section). To collect their knowledge and skills regarding the tasks (i.e., task dimensions, task requirements such as time, and task type), each participant was asked to fill out a questionnaire before and after each task. From the pre-task questionnaire we wanted to know about their knowledge about the task features before conducting the search. Post-task questionnaire solicited information regarding the overall experience with the task, their state of knowledge after completing the task. Table 1 describes the questions in the pre- and post-task questionnaires. Another important aspect is users' perceived difficulty of the search task because understanding of a task can affect users' search process and their interactions with the system (Bell & Ruthven, 2004). Therefore, the questions concerning task difficulty were adapted from Wu, Kelly, Edwards, and Arguello (2012).

4.3.2. Tasks characteristics and their representation

Participants completed three tasks consecutively during their sessions in the laboratory. They completed a warm-up task, task 1, and task 2 in that order. Descriptions of each task are provided below. Before each task, participants answered a pre-task questionnaire about their familiarity with the task. Participants were given 5 min, 20 min, and 20 min respectively to complete each task but could finish early. In the tasks, participants were asked to save bookmarked web pages and to construct reports based on their findings. The warm-up task also served to familiarize participants with the system and laboratory environment.

We designed each task by manipulating their level of complexity, the task goal, and the task product based on the task classification scheme proposed by Li and Belkin (2008). Li and Belkin (2008) identified various characteristics of a task and many of them

[1] https://www.techsmith.com/morae.html

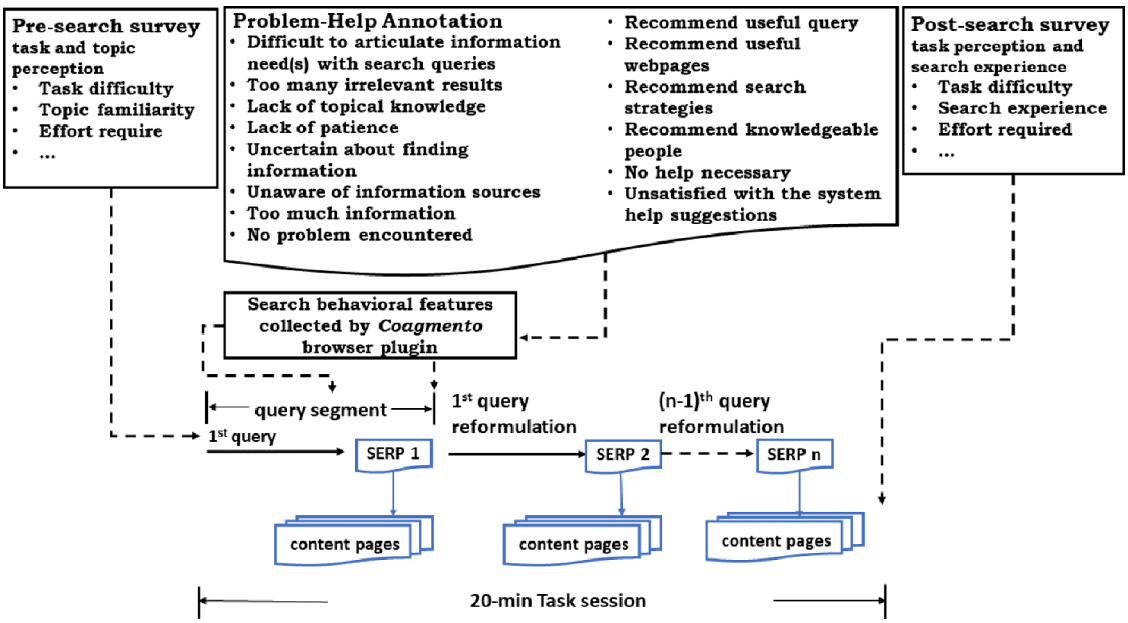

Fig. 2. Task session flow.

Table 1
Features of the tasks and users that were self-reported in a pre-task survey.

	Based on your reading of the task description, answer the questions carefully on 1 (Not at all) - 7 (Very familiar).
Acronyms	Feature
TFAM	How familiar are you with the topic of the task at this moment?
TEFF	How much effort will this task take?
	Based on your reading of the task description, please indicate your level of agreement with the following 10 statements on a 7-point scale from strongly disagree (1) to strongly agree (7).
Acronyms	Feature
TDIF	I think the task will be difficult.
TDET	The task description has a lot of details
TKSP	Right now, I know some specific things to look for to address the task
TUNP	There are aspects of the task, such as goal of the task, product of the task, scope of search, that are not specified in the description.
TNIN	The task description provides me with information that I did not already know.
UNOB	I know some specific types of useful information that I can not obtain from search engine.
KUINO	I know some specific types of useful information that I can obtain from search engine.
KQUS	Right now, I know the specific terms and queries that I should use to start my search.
GOSP	I think the task goal(s) are very specific.
NAIN	The information requested is narrowly focused.

were implied in our task design. Given the nature of search scenarios and study design, it was imperative that the tasks would be done by individuals ('Task doer' facet), be unique ('Time-Frequency' facet), be externally assigned ('Source of task' facet), and be done in a short time ('Time-Length' facet). Then, we carefully chose the aspects of product, goal, and objective task complexity. As mentioned in the previous section, we also asked questions about users' perception of task difficulty (Bell & Ruthven, 2004; Li & Belkin, 2008; Wu et al., 2012). Additionally, the tasks had different types of information needs. These binary labels were used in the subsequent task-based analysis. Tasks were defined to make the task situation relevant and realistic to the study participants, and to tailor the task to the particular group of participants (i.e., undergraduate students)(based on the simulated work tasks of Borlund (2003)). The task descriptions provided to participants are as follows:

Warm-up Task [Complexity = Moderate; Need = Cognitive; Product = Intellectual; Goal = Amorphous] (5 min): *You need to write a class report on HIV/AIDS treatments in Africa. For this, you need to answer a central question: what are the current available treatments of HIV/AIDS in China, Germany, USA, and Uganda?*

Task 1 [Complexity = High; Need = Cognitive; Product = Intellectual; Goal = Amorphous] (20 min): *Lara Dutta of India was crowned Miss Universe in 2000, and between 1994 and 2000 women from India won two Miss Universe competitions, four Miss World competitions, and many less well-known competitions. These facts inspired you to explore the relationship between these wins and the Indian government's decisions and policies in your final paper for Indian Society class. To what extent can decisions and policies of the Indian government be credited with these wins? As a part of your final paper, please offer your brief answer to this question and identify the useful pages (from the bookmarked pages) which were actually used for constructing your answer.*

Task 2 [Complexity = Low; Need = Social; Product = Intellectual; Goal = Amorphous] (20 min): *You will be attending a*

Table 2
The question for eliciting potential problems, and the possible responses. Acronyms in parentheses are used in future discussions.

Question: What problems are you facing at this moment? Select all that apply.
- (DIF) I do not know how to express my need in search queries.
- (IRR) I see a lot of not good or useless results.
- (TOP) I do not know enough about the topic.
- (PAT) I am feeling impatient.
- (CRE) I do not know if I can trust the information that I am seeing.
- (AWA) I may not know all the good or useful sources of information.
- (TMI) There is just too much information.
- (AVA) What I am looking for does not seem to be available.
- (None) No problem encountered.

social gathering this evening. It is a birthday party for your sister (a high school student) being held at a local restaurant. You do not know many of your sister's friends in attendance. You thought you could facilitate conversations with new friends if you were up-to-date on some recent topics of interest. You have decided to look into a wide expanse of topics and events (especially the topics you are not familiar with) based on your estimation of the other guests' interests, preferences and backgrounds. To be fully prepared, please create a list of interesting up-to-date topics/events (no less than 5 different topics). For each topic, please identify the useful web pages (from the bookmarked pages) and a very brief explanation for why you choose this as a potential topic suitable for your sister's birthday-party conversations.

4.3.3. Information seeking barriers and help

As mentioned in the previous section, the primary unit of analysis in this work is the perceived problems and helps at various points in the information seeking process. To elicit these from participants, we provided a questionnaire to them at the beginning of each query segment (before a query was actually formulated). The default URL of new browser tab was set as www.google.com, and participants had to open a new tab whenever they wanted to formulate a new query. Once a new tab (Google search home page) was opened, the browser automatically presented a pop-up window with a problem-help questionnaire. Participants had to report the problems they just encountered and the possible helps they would like to have in real-time search scenario on that questionnaire. By virtue of this design, it was possible to collect information about real-time search problem(s) and preferred possible help(s) before a user constructed a concrete, tentative solution (in this case, a query). To elicit useful, accurate responses regarding problems and helps from users, we adapted the relevant questions and concepts from existing research on gaps, barriers and helps in information seeking and sense-making (e.g., Chowdhury et al., 2011; Savolainen, 2015; Wang, Sarkar, & Shah, 2017; Xie & Cool, 2009). The problem and help questions included in the pop-up window are listed in Tables 2 and 3. We converged and combined different types of helps, barriers, and problems identified in above-mentioned studies and defined eight types of problem and four types of help which might provide insight into the problems they encounter, the strategies they use, and the patterns of interactions they engage in to resolve the problems. Note that the problems and preferred helps defined in the typologies are not mutually exclusive. Participants could select multiple options if applicable. However, if they selected "no problem encountered" or "no help needed," they could not select any other option simultaneously as other options would not be applicable in these situations. To avoid potential ordering effects, the order of options for all questions presented in the pop-up windows was randomized.

4.3.4. Implicit and explicit behavioral data

In addition, several features regarding the tasks, users, and their behaviors, described in Tables 4–6, were also collected. Each set of features is applied to address different research questions. For instance, previous segment features represent participants' behaviors immediately preceding the reporting of problem and help; these could represent possible indicators of their difficulties in finding information at the time. Perhaps a lack of bookmarks indicates that they saw useless results regarding their task and would prefer web pages with useful information as suggestions in order to fulfill their information need. Here, the previous query segment is relative to the current query segment at a given time-space situation within the search process. Likewise, session-based features are longer-term representations of prior behaviors. Current behaviors occur immediately after the reporting of problems and helps at a given point within the search process and could represent possible consequences of the possible problems and helps discovered at a current moment in the search process. For instance, "I am not satisfied with any help from the system" could entail a suite of behavioral changes, such as a lack of bookmarks, and could indicate the failure in obtaining useful information in an information seeking episode. Again, each set of features were used to address a different research question, as discussed below.

Table 3
The question for eliciting potential help, and the possible responses.

Question: What kind of things would help you at this moment? Select all that apply.
- (Query) Recommendations by the system about useful search queries.
- (Page) Recommendations by the system about potentially useful web pages.
- (Strategy) Recommendations about useful search steps and strategies.
- (People) Find me people who may be able to help.
- (Unsatisfied) I am not satisfied with any help from system, therefore, I would like to talk to someone whom I know (e.g., family, friends, colleagues).
- No help needed.

Table 4
Behavioral features for the previous query segment.

Acronym	Feature
PNCD	Number of distinct content pages over entire previous query segment
PNB	Number of bookmarks in previous segment
PTE	Time elapsed from the previous query
PTES	Time elapsed on SERPs in previous query
PQL	Query lengths over entire previous segment
PNQNC	Unique queries over previous segment without SERP clicks

Table 5
Behavioral features for the current query segment.

Acronym	Feature
CNB	Number of bookmarks in current segment
CTES	Time elapsed on SERPs in current query
CNCD	Number of distinct content pages over entire current query segment
CQL	Query lengths in current segment
CNQNC	Unique queries over current segment without SERP clicks

Table 6
Behavioral features accumulated over the session, up to the time of the reported problems and helps.

Acronym	Feature
STE	Time elapsed in whole session
STES	Time elapsed on SERPs in whole session
SNQD	Distinct queries over session
SNQNC	Distinct queries over session without SERP clicks
SNCD	Distinct content pages visited over session
SNCDPQ	Distinct content pages visited over session per unique query
SQLM	Mean of query lengths produced over entire session
SNB	Number of bookmarks over entire session

5. Data analysis and results

Data collected through the user study was analyzed using both descriptive and inferential methods. The unit of analysis was the problem and help situation at a given query segment when a participant was about to formulate a query. Operationally, the problem-help situation began when a user who was engaged in a information-searching task encountered a problem that prevented him or her from continuing the search task and marked the problem(s) and desired help(s) in the questionnaire, formulated a query, and ended when a user moved on in the search task by planning to formulate or reformulate another query. Although the query segments were recorded in the system log file, to ensure that the boundaries of the units of analysis were correctly isolated, we verified the beginning and end of each query segment by manually cleaning and coding the data set. After cleaning, the following analyses were performed based on the problem/help annotations collected from 208 query segments.

5.1. Descriptive analysis and results

Descriptive analyses were conducted to analyze the frequency of different problem-help situations. Summaries of the statistics of the problems and help are provided in Tables 7 and 8 and separated by task type. The total number of problems selected by participants was 357. The most prevalent problem was “I may not know all the good or useful sources of information” (78 times). The

Table 7
Counts and totals of perceived problems reported by participants in each task: DIF - Difficulty formulating queries, IRR - Irrelevant results, TOP - Lack of topic knowledge, PAT - Feeling impatient, CRE - Lack of credibility of information, AWA - Unaware of sources, TM0 - Too much information, AVA - Information source unavailable, None - No problem.

Task	Segments	DIF	IRR	TOP	PAT	CRE	AWA	TMI	AVA	Total	None
Warm-Up	53	6	8	20	11	3	20	9	3	80	15
1	84	37	31	33	22	13	40	13	25	214	13
2	71	6	8	3	6	8	18	12	2	63	31
Total	208	49	47	56	39	24	78	34	30	**357**	59

Table 8
Counts and totals of potential help reported by participants in each task: Page - Page recommendations by the system, People - People recommendations by the system, Query - Query recommendations by the system, Strategy - search strategy recommendations by the system, Unsatisfied - Unsatisfied with any help from system, None - No help needed.

Task	# Segments	Page	People	Query	Strategy	Total	Unsatisfied	None
Warm-Up	53	19	8	13	17	57	1	19
1	84	43	27	43	34	147	5	18
2	71	11	3	12	5	31	17	31
Total	208	73	38	68	56	**235**	23	68

total number of helps that participants indicated they would like to have at a given time when they perceived a problem in the search process was 235. “Recommendations by the system about potentially useful web pages” was the most preferred help (73 times). At some moments during the search process, participants mentioned they were not perceiving any problem (59 times) or did not think that they needed any help (68 times).

In order to understand participants’ query reformulation strategies and their relationships with the problems and helps identified during a particular query segment, each query was annotated based on the coding schema on query reformulation by Rha et al. (2016) where they proposed seven types of query reformulations – *Generalization (1), Specialization (2), Word Substitution (3), Repeat (4), New (5), Spelling Correction (6), and Stem Identical (7)*. In this study, the “New” query was the most commonly used reformulation type. In other words, most of the times participants created new queries using entirely new terms than the previous queries (Rha et al., 2016). The second most frequently used reformulation type was “Word Substitution”. Many times participants tended to substitute a portion of the query with new terms while keeping the length and some previous terms the same from the previous query. Reformulation patterns also differed between tasks. Participants applied “New” query reformulations to start a new search in both tasks 1 and 2. For task 1, they applied different formulation types several times. Figs. 3 and 4 show the distribution of query reformulations types of the whole sample.

On the contrary, there was not much difference in participants’ perceptions of the tasks. From the pre-task questionnaire, it seemed that the majority of the participants (12) perceived task 1 as the most difficult task, and the warm-up task was the most straightforward task to perform compared to the other two tasks (11). On the question of whether they were familiar with the task topics, all participants reported that they had no prior knowledge or familiarity with the topic in task 1, and the majority of them (11)

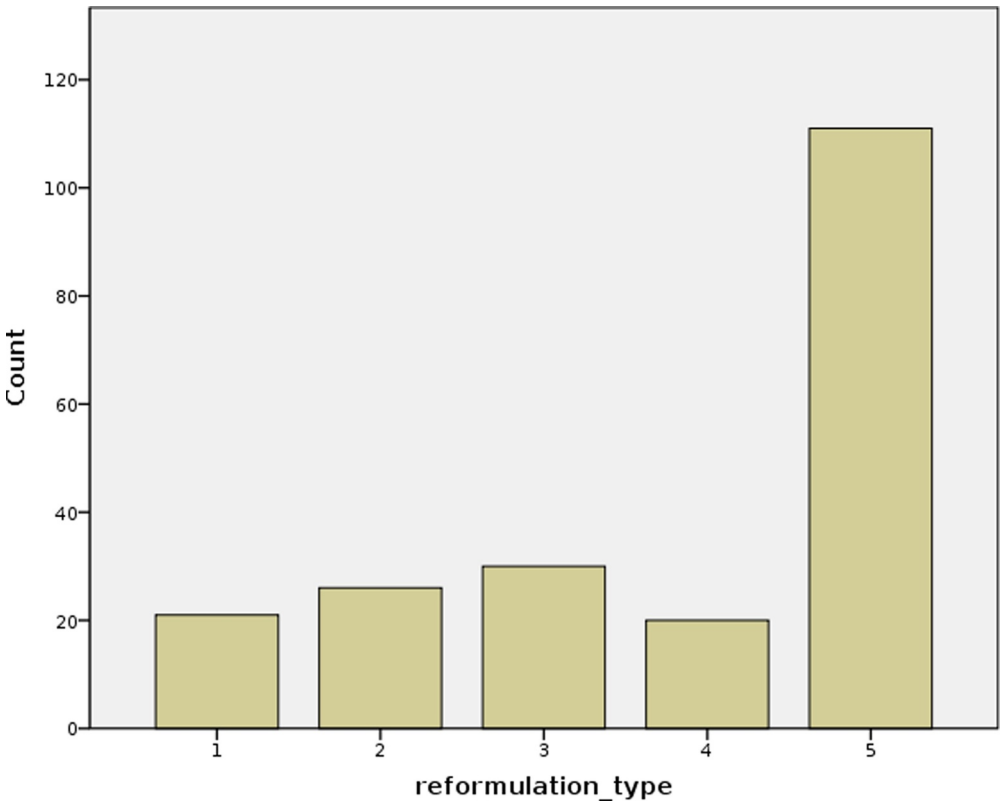

Fig. 3. Frequencies of query reformulations - (1) Generalization, (2) Specialization, (3) Word Substitution, (4) Repeat, (5) New.

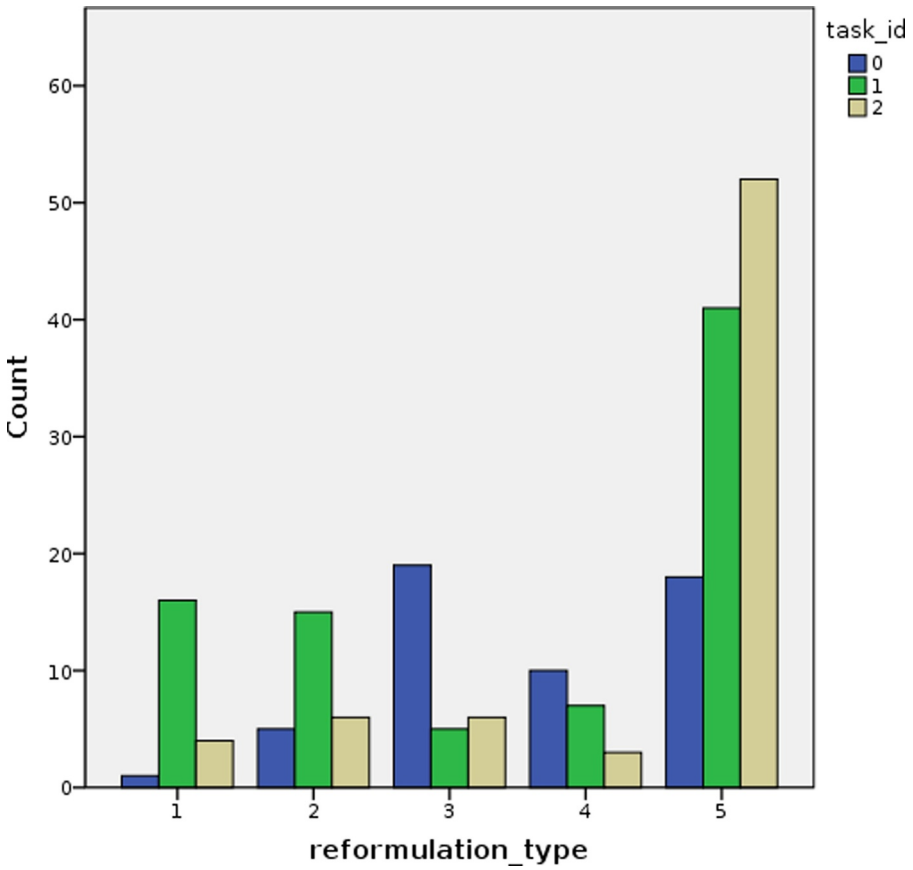

Fig. 4. Frequencies of query reformulations throughout in each task session - (1) Generalization, (2) Specialization, (3) Word Substitution, (4) Repeat, (5) New for the warm-up (0), Task 1 (1), Task (2).

were familiar with the topic of task 2. Similarly, participants reported that they had to exhibit much more effort and do broader explorations in task 1. Even after finishing the tasks, their perceptions of the tasks remained mostly unchanged.

5.2. Inferential analysis and results

The following sub-sections report the results of data analyses corresponding to the proposed research questions.

5.2.1. Predicting perceived problem from search behaviors

This subsection explores (1) how current, immediate past, and entire information search segment and associated seeking behaviors could reveal participants' overall understanding of the situational uncertainty (expressed as problems) at the subsequent (current) moment of search and (2) to what extent participants' information needs at current/a given point of time were observed through their subsequent searching behaviors. To this end, multiple binomial logistic regressions models were built for each anticipated problem or barrier reported by the participants in stages of information seeking. As the dependent variable - the problem was defined as a dichotomous variable with identified (1) or not identified (0), and we applied logistic regression to explore the relationships between the identified problems and various behavioral features from previous, current and entire session as independent variables. The outcomes and goodness of fit of the models are listed in the Tables 9–10. The statistically significant relations ($p < 0.05, p < 0.01, p < 0.001$) are marked with asterisks on the Table 9. The reported outputs of the regressions models can help researchers to explain the effects and relationships of these behavioral signals on the anticipated problems and participants' perceived information needs throughout different query segments. For example, the model informs that for every one unit increase in PTES (Time elapsed on SERPs in previous query), the log-odds of the problem of query articulation increases by 0.02.

According to the results presented in Table 9, the difficulty in articulating the current information need had statistically significant association with several behavioral measures, such as the amount of time spent on SERPs in the previous and current query segments, the number of distinct queries without clicks throughout the sessions, and negative relationship with number of bookmarks of web pages throughout the sessions. This result indicates that the time participants spent on the visualization of information sources presented by the information retrieval systems could generate a cognitive need to articulate their thoughts concisely in order to find more useful information eventually. Another possibility is that when participants were actively interacting with search pages of

Table 9
Coefficients and significance values for each variable in each regression (Significance: * = .05,** = .01, *** = .001). DIF - difficulty formulating query, IRR - irrelevant results, TOP - lack of topic knowledge, PAT - feeling impatient, CRE - lack of credibility of information, AWA - unaware of sources, TMI - too much information, AVA information source unavailable, None - no problem encountered, PNCD no. of distinct content pages over entire previous query segment, PNB no. of bookmarks in previous segment, PTE time elapsed from the previous query, PTES time elapsed on SERPs in previous query, PQL query lengths over entire previous segment, PNQNC unique queries over previous segment without SERP clicks, CNB no. of bookmarks in current segment, CTES - time elapsed on SERPs in current query, CNCD no. of distinct content pages over entire current query segment, CQL - query lengths in current segment, CNQNC - unique queries over current segment without SERP clicks, STE - time elapsed in whole session, STES -time elapsed on SERPs in whole session, SNQD - distinct queries over session, SNQNC - distinct queries over session without SERP clicks, SNCD - distinct content pages visited over session, SNCDPQ - distinct content pages visited over session per unique query, SQLM - mean of query lengths produced over entire session, SNB no. of bookmarks over entire session.

Features	DIF	IRR	TOP	IMP	CRE	AWA	TMI	AVA	None
Previous Segment Behavior Features									
PNCD	−0.11	−0.24	0.05	**−0.05***	0.12	0.09	−0.15	−0.06	**−0.03***
PNB	−0.15	−0.17	−0.18	−0.19	**−1.17***	−0.16	−0.23	−0.72	0.28
PTE	0.00	−0.00	0.00	0.00	−0.00	0.00	0.00	0.00	0.00
PTES	**0.02*****	**0.01****	−0.00	**0.01***	−0.00	0.00	0.00	**0.01****	−0.00
PQL	−0.09	**0.34*****	0.01	**0.22***	0.16	0.00	0.10	0.16	**−0.18***
PNQNC	0.23	−0.00	−0.71	−0.46	−0.23	−0.02	**−1.73****	−0.39	0.46
Current Segment Behavior Features									
CNB	−0.22	−0.01	−0.23	−0.19	**−1.33***	−0.22	−0.30	−0.12	**0.23***
CTES	**0.02*****	**0.01****	0.00	0.00	0.009	**0.01***	0.00	**0.02****	−0.00
CNCD	−0.21	−0.39	−0.19	−0.10	−0.17	−0.11	0.07	−0.62	0.16
CQL	−0.10	**0.18***	0.02	0.12	−0.00	−0.07	0.05	0.12	−0.02
CNQNC	−0.07	−0.33	−0.74	−0.69	−0.53	0.06	0.03	−1.05	0.10
Session Behaviors Features									
STE	0.00	−0.00	0.00	0.00	0.00	0.00	−0.00	−0.00	**0.003***
STES	0.00	**0.009*****	0.00	0.00	0.00	0.00	0.00	**0.01*****	−0.00
SNQD	−0.13	0.10	−0.14	0.21	0.19	−0.28	0.34	0.21	−0.20
SNQNC	**0.60****	0.04	−0.23	−0.15	0.07	0.33	**−0.66***	−0.04	−0.14
SNCD	−0.10	**−0.27***	−0.01	−0.08	−0.09	0.04	0.08	−0.16	0.02
SNCDPQ	0.00	**0.02***	−0.01	0.00	0.01	0.00	−0.00	0.01	−0.01
SQLM	−0.01	**0.01***	0.00	**0.01***	0.00	−0.00	0.01	0.01	−0.00
SNB	**−0.23***	0.00	−0.04	−0.15	**−0.79****	−0.08	**−0.34***	**−0.28***	0.10

Table 10
Logistic Regression: Goodness of Fit (χ^2 scores. Significance: * = .05,** = .01, *** = .001). DIF - difficulty formulating query, IRR - irrelevant results, TOP - lack of topic knowledge, PAT - feeling impatient, CRE - lack of credibility of information, AWA - unaware of sources, TMI - too much information, AVA information source unavailable, None - no problem encountered.

Features	df	DIF	IRR	TOP	PAT	CRE	AWA	TMI	AVA	None
Prev	6	18.07**	36.11**	7.79	20.16**	12.01	4.3	15.59*	24.79**	17.29**
Curr	5	19.82**	21.32**	8.17	7.23	14.07*	10.93	3.37	16.00**	8.17
Session	8	36.37**	35.90**	12.49	26.60**	20.86**	13.68	26.22**	34.81**	15.20

different types (e.g., longer dwell time), it is very likely that they had abstract information need or encountered an anomalous state of knowledge (cf. Belkin, 1980).

Besides, there are also statistically significant positive associations between the problem of not seeing a lot of good/useful results with users' dwell time on SERPs in previous, current and entire sessions, increasing query length in previous and current search segments, retrieved distinct web pages per query, and negative association with amount of retrieved distinct content pages throughout sessions. This result demonstrates that when participants did not find any useful or relevant information in the previous search results or perceived that the system was unable to provide the information they were looking for, they often took extra time and efforts on query formulation (especially on formulating long queries) so that more useful information could be successfully retrieved. The results also point to a negative relationship between the feeling of impatience and receiving distinct content pages in query segment, meaning that as search session proceeded and participants were not getting unique contents in the previous segment, they became more and more impatient on search interaction. If a participant experienced difficulty in expressing a query, had found too many irrelevant results, or was unaware of good sources of information, they tended to spend significantly more time browsing subsequent SERPs. Additionally, if a person distrusts information, then they will bookmark significantly less in the subsequent segment (9). If they felt that no problem was encountered, they would bookmark more content pages in the subsequent query segments.

Surprisingly, at the whole task session level, no browsing behavior feature (prior or current) was significantly associated with a perceived lack of topic knowledge or not knowing all of the good or useful sources. Perhaps these perceptions were more closely related to the prior context of the session and change little throughout the middle of a single search session (9). Nevertheless, not

Table 11
Weights (and significance) of self-reported task and user characteristics in regressions against reported problems, DIF - difficulty formulating query, IRR - irrelevant results, TOP - lack of topic knowledge, PAT - feeling impatient, CRE - lack of credibility of information, AWA - unaware of sources, TMI - too much information, AVA information source unavailable, None - no problem encountered, TDIF - task difficulty, TFAM - topic familiarity, TDET - detailed task description, TKSP - known information to address the task, TEFF - effort required, TUNP - missing information in the task description, TNIN - unknown information provided by the task description, UNOB - known useful information cannot be obtained from search engine, KUINO - known useful information, KQUS - known query to initiate search, GOSP - specific task goal(s), NAIN - requested information is narrowly focused.

Features	DIF	IRR	TOP	PAT	CRE	AWA	TMI	AVA	None
TDIF	0.41	**0.65****	0.04	**0.54****	0.45	0.30	−0.07	0.04	−0.19
TFAM	−0.03	−0.18	**−0.44***	−0.23	**0.85*****	0.09	0.25	−0.37	**−0.28***
TDET	**0.56***	−0.28	0.26	0.34	0.14	0.08	**0.43****	0.19	−0.03
TKSP	−0.33	0.18	0.22	−0.09	0.30	0.23	−0.00	0.59	−0.35
TEFF	**−0.59***	−0.63	0.28	**−0.72***	0.21	−0.11	−0.14	−0.35	0.11
TUNP	−0.03	−0.01	0.12	−0.18	0.12	0.03	0.09	−0.09	−0.07
TNIN	0.35	**0.42***	**−0.50****	−0.00	0.26	**0.26***	−0.12	0.36	**−0.34****
UNOB	−0.26	**−0.37***	**−0.39****	**−0.57****	−0.09	−0.11	**−0.36****	−0.01	**0.32***
KUINO	−0.32	−0.10	−0.04	0.43	−0.03	−0.07	0.11	**−0.82****	−0.13
KQUS	0.31	−0.21	−0.45	−0.51	−0.50	−0.02	−0.20	−0.42	**0.57***
GOSP	−0.35	**1.04****	−0.34	0.07	−0.47	−0.39	−0.28	0.79	0.26
NAIN	0.16	**−0.70***	0.45	−0.28	0.70	0.20	0.00	−0.57	−0.28

knowing how to express one's need, seeing a lot of good or useless results, and the inability to find available information were associated with changes in all three measured components. This result suggests that perceptions of problems not only arise from the previous strategies employed over the immediate segment (e.g., no clicks on SERPs) but emerge from the user's recognition of general patterns of their general search behavior over the session (e.g., few bookmarks) (9). Additionally, the perception and recognition of problems can result in significant changes in Web search behavior (e.g., increased time spent on future SERPs) (9).

The goodness of fit of the models reported in the Table 10 shows how the model as a whole is statistically significant. It shows how much the predicted frequency and observed frequency match closely with each other, and that the more closely they match, the better the fit. The table shows the models to predict perceived problems of articulation of queries, useless and irrelevant information, and unavailability of information can perform statistically significantly better to foresee these problems through users' previous, current and session behaviors. Furthermore, behaviors at an earlier search session and throughout the whole session can perform better in predicting the lack of patience and the problem with too much information (9).

As mentioned earlier, users use their knowledge, skills, and experience to achieve their goals however, their existing knowledge are not always sufficient to deal with all problematic situations that may arise. Therefore, we also examined to what extent perceived problems were dependent on the users' states of knowledge or task features reported by the users by building binomial logistic regression models with problems as dependent variables, and users' characteristics and perceived tasks' characteristics as independent variables. Table 11 reports results from the models and Table 12 reports the goodness of fit of the models. According to the report, the perceived problem of visualizing too many useless and irrelevant results is positive and statistically significant to user perceived task difficulty which indicates that when effects of task difficulty increase over time in the search process, the perceived problem of seeing irrelevant information also increases. Moreover, the goal specificity, new information in task descriptions, and participants not knowing useful information also affected the perceived problem of irrelevant results.

As the task difficulty perceived by the participants increased over time as the search proceeded, participants would run out of patience in the process (11). The task familiarity negatively affected the trustworthiness of retrieved information. As the session proceeded, the familiarity of the tasks increased, and participants gained more knowledge about a particular topic, could make a more informed quality judgment, and that decreased their trust towards information. Contrary to common beliefs, when participants thought that the task descriptions were detailed, that had a positive statistically significant relationship with problems with articulation of needs and finding too much information (9).

5.2.2. Predicting helps from behaviors and problems

A complementary component to a searcher's perceived problems is the help a person perceives as best to address their information need. To examine how a person's problematic situation is related to the perceived best help for that person, we constructed a classification problem, using query segments annotated with problems and help which had some prior behaviors and therefore not

Table 12
Logistic Regression - Task-user characteristics: Goodness of Fit (χ^2 scores. Significance: *=.05,**=.01, ***=.001). DIF - difficulty formulating query, IRR - irrelevant results, TOP - lack of topic knowledge, PAT - feeling impatient, CRE - lack of credibility of information, AWA - unaware of sources, TMI - too much information, AVA information source unavailable, None - no problem encountered.

Features	df	DIF	IRR	TOP	PAT	CRE	AWA	TMI	AVA	None
Task-user	12	61.21***	49.24***	60.32***	33.74***	31.78**	25.38*	16.66	44.31***	32.66

Table 13
Accuracy scores for predicting the currently reported help using previous query segment features, session features, the currently reported problem, and all features combined: Page - Page recommendations by the system, People - People recommendations by the system, Query - Query recommendations by the system, Strategy - search strategy recommendations by the system, Unsatisfied - Unsatisfied with any help from system, None - No help needed.

Features	No Help	Page	People	Query	Strategy	Unsatisfactory
Previous Segment	0.53	0.64	**0.78**	0.67	0.72	0.94
Session	0.61	0.69	**0.78**	0.67	0.75	**0.97****
Problems	**0.75****	0.61	0.75	0.72	0.69	0.92
All	**0.75****	0.61	0.75	**0.78**	0.67	0.94
MFreq	0.61	**0.72****	0.78	0.78	**0.81****	0.92
Random	0.53	0.5	0.61	0.61	0.64	0.92
Always-Yes	0.39	0.28	0.22	0.22	0.19	0.08

the first segment in a session (178). We examined which aspects of a searcher's session most contributed to accurate classification of a person's perceived best help at a given moment. Considered aspects include behavior in the immediately prior query segment, behavior over the entire session, and even the reported problems (under the assumption that these can be accurately determined). More accurate classifications help indicate the extent to which these components contribute to the perceived best help of a person at a given moment. We applied a simple logistic regression classifier, with an 80/20 split on training/test data. This was compared to three simple baselines: a random baseline, a most frequent labeling baseline (which always labeled a problem as absent), and a baseline that always labels a problem as present. The features used are the same as those in Tables 4–6. An overview of the results is provided in Tables 13–16.

We apply the classic machine learning definitions of precision, recall, and F1 here, which are given in the below equations. The following definitions assume that a "positive" is when the perceived problem/help is present (i.e., reported by the user or predicted by the system). A "true positive" occurs when the system predicts that a perceived problem/help was present when it was also in fact reported by the user. A "false positive" occurs when the system predicts the problem/help was present when the user did not report it. Similar definitions can be formulated for "true negative" and "false negative", where the system predicts that the user did not experience the problem or desired help.

$$Precision = \frac{\#\text{true positives}}{\#\text{predicted positives by algorithm}} \tag{1}$$

$$Recall = \frac{\#\text{true positives}}{\#\text{real positives in data}} \tag{2}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3}$$

Overall, search characteristics can be seen as influencing the perceived best help at a given moment in the search. Accuracy scores would seem to dictate otherwise. Naive baselines (namely, the most frequent labeler) tend to have comparable or significantly better accuracy and would suggest that the selected behaviors cannot be used to predict perceived help. However, this is due to the skewed nature of the problem - a single perceived help does not occur the majority of the time during sessions. Precision scores show whether a classifier would be correct when flagging that a person has a perceived need. Table 14 shows that constructing such a classifier is possible using the selected behaviors from the current study. Namely, whether no help is necessary can be determined by the number of problems and all combined browsing features. Session features can be used to determine that a page is the desired help. The previous segment or the session can be used to determine that a person is the desired help. All of the features combined can be used to determine that a query is the desired help. The entire prior session can be used to determine whether a strategy is the desired help. Dissatisfaction with the search system can be determined by the entire session. F1 scores broadly give further validation to these

Table 14
Precision scores for predicting the currently reported help using previous query segment features, session features, the currently reported problem, and all features combined: Page - Page recommendations by the system, People - People recommendations by the system, Query - Query recommendations by the system, Strategy - search strategy recommendations by the system, Unsatisfied - Unsatisfied with any help from system, None - No help needed.

Features	No Help	Page	People	Query	Strategy	Unsatisfactory
Previous Segment	0	0.33	**0.5****	0.3	0.2	0.67
Session	0.5	**0.46****	**0.5****	0.25	**0.42****	**1****
Problems	**0.78****	0.33	0.33	0.42	0.25	0
All	**0.78****	0.33	0.33	**0.5****	0.27	0.67
MFreq	0	0	0	0	0	0
Random	0.33	0.1	0	0.2	0.2	0.5
Always-Yes	0.39	0.28	0.22	0.22	0.19	0.08

Table 15
Recall scores for predicting the currently reported help using previous query segment features, session features, the currently reported problem, and all features combined: Page - Page recommendations by the system, People - People recommendations by the system, Query - Query recommendations by the system, Strategy - search strategy recommendations by the system, Unsatisfied - Unsatisfied with any help from system, None - No help needed.

Features	No help	Page	People	Query	Strategy	Unsatisfactory
Previous Segment	0	0.3	0.12	0.38	0.14	0.67
Session	0.21	0.6	0.12	0.25	0.71	0.67
Problems	0.5	0.4	0.12	0.62	0.29	0
All	0.5	0.4	0.12	0.75	0.43	0.67
MFreq	0	0	0	0	0	0
Random	0.21	0.1	0	0.25	0.29	0.67
Always-Yes	**1****	**1****	**1****	**1****	**1****	**1****

Table 16
F1 scores for predicting the currently reported help using previous query segment features, session features, the currently reported problem, and all features combined: Page - Page recommendations by the system, People - People recommendations by the system, Query - Query recommendations by the system, Strategy - search strategy recommendations by the system, Unsatisfied - Unsatisfied with any help from system, None - No help needed.

Features	No help	Page	People	Query	Strategy	Unsatisfactory
Previous Segment	0	0.32	0.2	0.33	0.17	0.67
Session	0.3	**0.52****	0.2	0.25	**0.53****	**0.8****
Problems	**0.61****	0.36	0.18	0.5	0.27	0
All	**0.61****	0.36	0.18	**0.6****	0.33	0.67
MFreq	0	0	0	0	0	0
Random	0.26	0.1	0	0.22	0.24	0.57
Always-Yes	0.56	0.43	**0.36****	0.36	0.33	0.15

precision scores.

The specific details of these results should be taken with a grain of salt. This analysis only scratches the surface regarding the extent to which the desired help of a person can be predicted at a given moment, using a very modest classifier. Future work would need to determine the upper limits of classification: using a richer suite of browsing behaviors and richer machine learning techniques. However, it is possible to draw some initial conclusions from the results. First, for several desired helps, the most reliable indicators are behaviors over the entire session - specifically for pages, people, and strategies, and general dissatisfaction with the system (Table 14). For indicating whether queries are desired or whether no help is necessary, the problems experienced at a given moment are generally good indicators. However, we came to the counterintuitive conclusion that reported problems are not really predicative of the type of the help(s) desired at the moment. For instance, “I see a lot of not good or useless results” could perhaps be a strong and reliable indicator of the desire for a useful web page. Yet this is not the case. The desired help at a given moment is more than just a function of the problems experienced at hand. Is it also a function of a user’s knowledge? Of their affective states at a given moment? Future work would need to delve deeper into this.

6. Discussion and implications

This study explores the connection between users’ implicit and explicit searching behaviors, the constraints and barriers people perceived while looking for information online, and preferred help to mitigate those barriers in different stages of an information search episode. From the data analysis, some interesting observations emerge. Some of the observations are evident in existing studies and are subject to common sense. However, some subtle observations additionally emerged from the results that can be important for understanding barriers that individuals encounter in different stages of search process, his/her realizations of information needs, the help they prefer, and search behaviors.

This study reveals that there are statistically significant associations between individuals’ perceived problems and preferred helps in search process, and these relationships can be explained through the search behaviors. Therefore, it is possible to predict users’ perceived problems in different stages of information search process by mapping users’ implicit and explicit search behaviors. The amount of time spent on previous search results could be an indicator of the potential problems of articulation of needs into queries, perceiving useless results, and not getting useful sources in the subsequent search stage. The possible explanation behind these observations could be that the user was having trouble finding the information he/she was looking for in the search results. Another possible explanation could be that since the previous query retrieved a large number of irrelevant web pages, the user wanted to spend some time to formulate the query so that it could retrieve relevant information to the query.

Other features of previous search stage, such as number of distinct content pages, unique queries, length of query, and number of bookmarks could predict the subsequent perception of barriers in users’ minds – “I see a lot of not good or useless results”, “I am feeling impatient”, “I do not know if I can trust the information that I am seeing”, “There is just too much information”, “What I am

looking for does not seem to be available." By observing current search behaviors, it is possible to infer the presence of these problems – "I do not know how to express my need in search queries", "I see a lot of not good or useless results", "I do not know if I can trust the information that I am seeing", "I may not know all the good or useful sources of information", "What I am looking for does not seem to be available". The study also suggests that the perceptions of problems not only arise from the previous strategies employed over the immediate segment (e.g., no clicks on SERPs) but also emerge from users' perception and recognition of general patterns of their general search behavior over the session (e.g., few bookmarks). These perceived problems in search can result in significant behavioral changes in following query segments (e.g., increased time spent on future SERPs).

The study also indicates that various behaviors over a search session can serve as an indicator of what helps are desired at a given moment. However, the desired help at the moment is not merely a function of the perceived problems faced that moment. Sometimes the perceived problems are a significant contributor. However, for instance, a desire for a page recommendation or a strategy recommendation seems to be primarily governed by behaviors accumulated over the entire search session. Therefore, the perceived best help can be seen as not only a function of the problems they perceive but as a function of the strategies they employ over the session. Therefore, help can be seen as a function of their information needs, the strategies they use to meet those needs, and the problems users encounter.

Additionally, the study also found that users' perception of the task difficulty, the description of the tasks, and knowledge and familiarity with similar tasks also served as indicative factors for perceived problems and preferred help. Task and user characteristic could be an influencing factor for search behaviors. We saw that task 1 with high complexity significantly influenced the behaviors, and perceived problems and helps. Behaviors such as reformulation of a query for the subsequent search were influenced by task type. For the social task with low complexity, users mostly searched with an entirely new query most of the time, whereas, for cognitive and moderate to high complexity tasks, users used new and substituted queries as well. In the pre-task questionnaire, the majority of the users perceived task 1 as the most challenging task, and possibly for that perception, the frequency of each preferred help is higher for task 1.

Finally, the study seeks to obtain a more comprehensive, quantifiable understanding of the information need that people realize at a given moment by conceptually mapping and empirically measuring these three aspects in a information search process. The findings show that by combining the perceived problems, help and search behaviors at a given moment, it is possible to infer users' momentary need. For example, if at a given moment of the search process, when a user realized his or her lack of trust on information retrieved by the system and asked for useful pages, and at the same time the number of bookmarks pages decreased, these could have been an indicator of his or her need for more useful information.

7. Conclusions

Information need is a widely researched concept within information science. However, despite its importance, there is still a lack of understanding of what information need is and what its dimensions are. Therefore, there is a need to go beyond the approaches used thus far, especially to conceptualize as well as operationalize the construct of information need. The purpose of this study is to propose a new methodology to examine the concept of information need in different stages of the information search process by analyzing the relationship between searchers' perceived problems and the preferred helps through their behaviors.

However, this experimental study has some limitations. First, we used simulated task scenarios in a controlled laboratory setting that did not arise from participants' daily lives and actual needs. These might have triggered some of the participants to complete the tasks differently and perceive different types of problems and helps at a given moment. Dissatisfied with the information provided by the system, one of our participants mentioned that in real life, if they had to do these tasks, they would have contacted an expert or a family member and asked for help to solve a particular problem or query suggestion, but could not do so due to the study design. Second, the participants had to select potential problems and possible helps from a list of pre-selected problems and helps which might not truly reflect and represent users' actual problems or help needs. Although we tried to include all possible problems and helps related to the Web search environment, it was possible that the problem or the help a user faced was not in the list. Participants may have felt that they had to select something from the given choices; thus, participants' reports might not accurately depict the actual problem or help at every moment of search interaction. Third, the controlled laboratory setting and time constraint might also influence participants' responses. Lastly, Some of the measures employed in this study were based on participants self-reported data. As with other self-reported data research, it may prone to some subjective biases.

Irrespective of the limitations, one of the main contributions of this study is that it provides a detailed procedure to identify and observe information need in different stages of an information search process. Another important implication of this study is that it presents a way to identify and understand barriers (reported problems) that people encounter while searching for information using automated information retrieval systems and try to mitigate the problems (by looking for help) in subsequent stages through their previous search behaviors. Thus, the study examined the relationship between behaviors, perceived problems and helps at independent moments in a session. It has long been known that search sessions do not consist of isolated moments but an interleaved series of strategies. Extensions to this work inevitably involve additional data collection. This could include studying the nature of a sequence of problems and helps (therefore a sequence of needs) and determining problems' and helps' relationships with broader patterns of strategies. Related extensions include closer examinations of the importance of domain or topic expertise, as well as expanding the task types that are studied, whether general Web tasks or even discipline-specific tasks (e.g., patent search).

Furthermore, when searching for information online, a person will leave a trail of queries and document views. Given this history and even personal preferences, a search engine may recommend related queries or potentially useful documents when the next query is issued. More broadly, it may be more useful to recommend a strategy or a process, not only providing a set of results but a process

by which to reach them for future tasks. A useful analogy to this system-user relationship is that of a teacher-student relationship, in which the teacher illuminates how the student can best resolve future problems that are similar. Thus, this study also presents a predictive model of users' process of sense-making of a problematic situation while using Web-based information search system based on implicit behavioral measures and explicit feedback, thus contributing to the improvement of user search experience and retrieval systems. To effectively assist people in their information searching especially in Web environment, we first need to identify the types of problems people experience for which they require help and to better understand the nature of problematic situations, gaps or needs. Moreover, as the study also examines the effects of several contextual factors in the context of search (e.g., users' search experiences, domain knowledge, the perception of task complexity, and social and behavioral characteristics), it has implications for designing personalized search support and system recommendations for users from different populations and literacy levels.

Acknowledgments

The study reported here was funded by the NSF grant #IIS-1717488. Thus, we would like to acknowledge the NSF for their support, our study participants for their participation, and the anonymous reviewers for their helpful comments and suggestions.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.ipm.2019.102069 .

References

Afzal, W. (2017). A proposed methodology for the conceptualization, operationalization, and empirical validation of the concept of information need. *Information Research: An International Electronic Journal, 22*(3).

Al-Fedaghi, S. S. (2008). *Integration of information needs and seeking. 2008 IEEE International conference on information reuse and integration, IEEE IRI-2008*473–478.

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

Bates, M. J. (2002). Toward an integrated model of information seeking and searching. *The New Review of Information Behaviour Research, 3*, 1–15.

Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science, 5*(1), 133–143.

Belkin, N. J., Oddy, R. N., & Brooks, H. M. (Oddy, Brooks, 1982a). ASK for information retrieval: Part I. Background and theory. *Journal of Documentation, 38*(2), 61–71.

Belkin, N. J., Oddy, R. N., & Brooks, H. M. (Oddy, Brooks, 1982b). ASK for information retrieval: Part II. Results of a design study. *Journal of Documentation, 38*(3), 145–164.

Belkin, N. J., & Vickery, A. (1985). Interaction in information systems: A review of research from document retrieval to knowledge-based systems. *Library and Information Research Report, 35*.

Bell, D. J., & Ruthven, I. (2004). *Searchers assessments of task complexity for web searching. European conference on information retrieval.* Springer57–71.

Borlund, P. (2003). The iir evaluation model: a framework for evaluation of interactive information retrieval systems. *Information Research. An International Electronic Journal, 8*(3).

Borlund, P. (2016). A study of the use of simulated work task situations in interactive information retrieval evaluations: A meta-evaluation. *Journal of Documentation, 72*(3), 394–413.

Borlund, P., & Ingwersen, P. (1997). The development of a method for the evaluation of interactive information retrieval systems. *Journal of Documentation, 53*(3), 225–250.

Broder, A. (2002). A taxonomy of web search. *ACM SIGIR Forum, 36*(2), 3. https://doi.org/10.1145/792550.792552.

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. https://doi.org/10.1002/asi.10064.

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

Capra, R., Arguello, J., O'Brien, H., Li, Y., & Choi, B. (2018). *The effects of manipulating task determinability on search behaviors and outcomes. The 41st international ACM SIGIR conference on research & development in information retrieval.* ACM445–454.

Looking for information: A survey of research on information seeking, needs, and behavior. In D. O. Case (Ed.). Bingley, UK: Emerald Group Publishing Limited.

Chowdhury, S., Gibb, F., & Landoni, M. (2011). Uncertainty in information seeking and retrieval: A study in an academic environment. *Information Processing & Management, 47*(2), 157–175.

Cole, C. (2011). A theory of information need for information retrieval that connects information to knowledge. *Journal of the American Society for Information Science and Technology, 62*(7), 1216–1231.

Cole, C., Leide, J., Beheshti, J., Large, A., & Brooks, M. (2005). Investigating the anomalous states of knowledge hypothesis in a real-life problem situation: A study of history and psychology undergraduates seeking information for a course essay. *Journal of the American Society for Information Science and Technology, 56*(14), 1544–1554. https://doi.org/10.1002/asi.20248.

Dervin, B. (1976). Strategies for dealing with human information needs: Information or communication? *Journal of Broadcasting & Electronic Media, 20*(3), 323–333.

Dervin, B. (1983a). Information as a user construct: The relevance of perceived information needs to synthesis and interpretation. In S. A. Ward, & L. J. Reed (Eds.). *Knowledge structure and use: Implications for synthesis and interpretation* (pp. 153–183). Philadelphia, PA: Temple University Press.

Dervin, B. (1983b). *An overview of sense-making research: Concepts, methods, and results to date.* Seattle: School of Communications, University of Washington.

Dervin, B. (1992). From the mind's eye of the user: The sense-making qualitative-quantitative methodology. In J. D. Glazier, & R. R. Powell (Eds.). *Qualitative research in information management* (pp. 61–84). Englewood, CO: Libraries Unlimited.

Dervin, B., Foreman-Wernet, L., & Lauterbach, E. (2003). *Sense-making methodology reader: Selected writings of Brenda Dervin.* Cresskill, NJ: Hampton Press.

Dervin, B., & Nilan, M. (1986). Information needs and uses. In M. E. Williams (Ed.). *Annual review of information science and technology (ARIST), vol. 21* (pp. 3–33). White Plains, NY: Knowledge Industry Publications. https://doi.org/10.1016/S0022-5371(63)80069-9.

Dumais, S. T., & Belkin, N. J. (2005). The trec interactive tracks: Putting the user into search. *TREC: Experiment and Evaluation in Information Retrieval,* 123–152.

Dutta, R. (2009). Information needs and information-seeking behavior in developing countries: A review of the research. *The International Information & Library Review, 41*(1), 44–51.

Feild, H. A., Allan, J., & Jones, R. (2010). *Predicting searcher frustration. Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval.* ACM34–41.

Fox, S., Karnawat, K., Mydland, M., Dumais, S., & White, T. (2005). Evaluating implicit measures to improve web search. *ACM Transactions on Information Systems (TOIS), 23*(2), 147–168.

Hoenkamp, E. (2009). *On the notion of "an information need". Conference on the theory of information retrieval*354–357.

Hoenkamp, E. C. (2015). *About the 'compromised information need' and optimal interaction as quality measure for search interfaces. Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval - SIGIR '15*835–838. https://doi.org/10.1145/2766462.2767800.

Ikoja-Odongo, R., & Ocholla, D. N. (2003). Information needs and information-seeking behavior of artisan fisher folk of uganda. *Library & Information Science Research, 25*(1), 89–105.
Jansen, B. J. (2006). Using temporal patterns of interactions to design effective automated searching assistance. *Communications of the ACM, 49*(4), 72–74.
Jansen, R. S.-Y., & B. J. (2010). The seventeen theoretical constructs of information searching and information retrieval. *Journal of the American Society for Information Science and technology, 61*(8), 1517–1534.
Jiang, J., He, D., & Allan, J. (2014). *Searching, browsing, and clicking in a search session: Changes in user behavior by task and over time. Proceedings of the 37th international ACM SIGIR conference on research & #38; Development in information retrievalSIGIR '14*ACM607–616.
Julien, H., & Duggan, L. J. (2000). A longitudinal analysis of the information needs and uses literature. *Library and Information Science Research, 22*(3), 291–309. https://doi.org/10.1016/S0740-8188(99)00057-2.
Kari, J. (1998). *Making sense of sense-making: From metatheory to substantive theory in the context of paranormal information seeking. Nordis-Net workshop:(Meta) theoretical stands in studying library and information institutions: individual, organizational and societal aspects, Oslo, Norway*. Citeseer1–20.
Kelly, D. (2009). Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends® in Information Retrieval, 3*(1–2), 1–224.
Kuhlthau, C., & Vakkari, P. (1999). Information seeking in context (ISIC). *Information Processing & Management, 35*(1999), 723–725.
Kuhlthau, C. C. (1991a). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science, 42*(5), 361–371.
Kuhlthau, C. C. (1991b). Inside the search process: seeking meaning from the users perspective. *Journal of the American Society for Information Science, 42*(5), 361–371.
Li, Y., & Belkin, N. J. (2008). A faceted approach to conceptualizing tasks in information seeking. *Information Processing & Management, 44*(6), 1822–1837.
Liu, J., & Belkin, N. J. (2010). *Personalizing information retrieval for multi-session tasks: The roles of task stage and task type. Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval*. ACM26–33.
Liu, J., Cole, M. J., Liu, C., Bierig, R., Gwizdka, J., Belkin, N. J., ... Zhang, X. (2010). *Search behaviors in different task types. Proceedings of the 10th annual joint conference on Digital libraries*. ACM69–78.
Liu, Y., Chen, Y., Tang, J., Sun, J., Zhang, M., Ma, S., & Zhu, X. (2015). *Different users, different opinions: Predicting search satisfaction with mouse movement information. Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*. ACM493–502.
Lundh, A. (2010). Studying information needs as question-negotiations in an educational context: a methodological comment. *Information Research: An International Electronic Journal, 15*(4), n4.
Mao, J., Liu, Y., Zhou, K., Nie, J.-Y., Song, J., Zhang, M., ... Luo, H. (2016). *When does relevance mean usefulness and user satisfaction in web search? Proceedings of the 39th international ACM SIGIR conference on research and development in information retrieval*. ACM463–472.
Marchionini, G. (1997). *Information seeking in electronic environments. 9*. Cambridge University Press.
Marchionini, G. (2010). Information concepts: From books to cyberspace identities. *Synthesis Lectures on Information Concepts, Retrieval, and Services, 2*(1), 1–105.
Marchionini, G., Dwiggins, S., Katz, A., & Lin, X. (1993). Information seeking in full-text end-user-oriented search systems: The roles of domain and search expertise. *Library & Information Science Research, 15*(1), 35–69.
Moshfeghi, Y., Triantafillou, P., & Pollick, F. E. (2016). *Understanding information need: An fmri study. Proceedings of the 39th international ACM SIGIR conference on research and development in information retrieval*. ACM335–344.
Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review, 106*(4), 643.
Reinhard, C. D., & Dervin, B. (2012). Comparing situated sense-making processes in virtual worlds: Application of dervins sense-making methodology to media reception situations. *Convergence, 18*(1), 27–48.
Rha, E. Y., Mitsui, M., Belkin, N. J., & Shah, C. (2016). *Exploring the relationships between search intentions and query reformulations. Proceedings of the 79th ASIS&T annual meeting: Creating knowledge, enhancing lives through information & technology*. American Society for Information Science48.
Ruthven, I. (2019a). The language of information need: Differentiating conscious and formalized information needs. *Information Processing and Management, 56*(1), 77–90. https://doi.org/10.1016/j.ipm.2018.09.005.
Ruthven, I. (2019b). *Making meaning: A focus for information interactions research. Proceedings of the 2019 conference on human information interaction and retrieval*. ACM163–171.
Saracevic, T. (1975). Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the Association for Information Science and Technology, 26*(6), 321–343.
Saracevic, T., & Kantor, P. (1988). A study of information seeking and retrieving. III. Searchers, searches, and overlap. *Journal of the American Society for information Science, 39*(3), 197–216.
Savolainen, R. (1995). Everyday life information seeking: Approaching information seeking in the context of "way of life". *Library & Information Science Research, 17*(3), 259–294.
Savolainen, R. (2012). Conceptualizing information need in context. *Information Research, 17*(4).
Savolainen, R. (2015). Cognitive barriers to information seeking: A conceptual analysis. *Journal of Information Science, 41*(5), 613–623. https://doi.org/10.1177/0165551515587850.
Savolainen, R. (2016). Approaches to socio-cultural barriers to information seeking. *Library & Information Science Research, 38*(1), 52–59.
Shah, C., & González-Ibáñez, R. (2011). *Evaluating the synergic effect of collaboration in information seeking. Proceedings of the 34th international ACM SIGIR conference on research and development in information retrieval*. ACM913–922.
Swigon, M. (2011). Information barriers in libraries: Types, typologies and polish empirical studies. *Library Management, 32*(6/7), 475–484. https://doi.org/10.1108/01435121111158619.
Talja, S., & Nyce, J. M. (2015). The problem with problematic situations: Differences between practices, tasks, and situations as units of analysis. *Library and Information Science Research, 37*(1), 61–67. https://doi.org/10.1016/j.lisr.2014.06.005.
Taylor, R. S. (1962). The process of asking questions. *Journal of Documentation, 13*(4), 391–395. https://doi.org/10.1002/asi.5090130405.
Taylor, R. S. (1968). Question-negotiation and information seeking in libraries. *College & Research Libraries, 29*(3), 178–194.
Vakkari, P. (1999). Task complexity, problem structure and information actions: Integrating studies on information seeking and retrieval. *Information Processing & Management, 35*(6), 819–837.
Wang, Y., Sarkar, S., & Shah, C. (2017). *Investigating information seekers' selection of interpersonal and impersonal sources. Proceedings of the 2017 conference on conference human information interaction and retrieval*. ACM353–356.
White, R. W., & Huang, J. (2010). *Assessing the scenic route: measuring the value of search trails in web logs. Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval*. ACM587–594.
White, R. W., Jose, J. M., & Ruthven, I. (Jose, Ruthven, 2003a). *An approach for implicitly detecting information needs. Proceedings of the twelfth international conference on information and knowledge management*. ACM504–507.
White, R. W., Jose, J. M., & Ruthven, I. (Jose, Ruthven, 2003b). *An approach for implicitly detecting information needs. Proceedings of the twelfth international conference on information and knowledge management - CIKM '03*504. https://doi.org/10.1145/956863.956959.
White, R. W., Jose, J. M., & Ruthven, I. (2006). An implicit feedback approach for interactive information retrieval. *Information Processing & Management, 42*(1), 166–190.
Wildemuth, B., Freund, L., & Toms, E. G. (2014). Untangling search task complexity and difficulty in the context of interactive information retrieval studies. *Journal of Documentation, 70*(6), 1118–1140. https://doi.org/10.1108/JD-03-2014-0056.
Wilson, T. D. (1981). On user studies and information needs. *Journal of Documentation, 37*(1), 3–15.
Wu, W.-C., Kelly, D., Edwards, A., & Arguello, J. (2012). *Grannies, tanning beds, tattoos and nascar: Evaluation of search tasks with varying levels of cognitive complexity. Proceedings of the 4th information interaction in context symposium*. ACM254–257.
Xie, I., & Cool, C. (2009). Understanding help seeking within the context of searching digital libraries. *Journal of the American Society for Information Science and Technology, 60*(3), 477–494.