Identifying and Predicting the States of Complex Search Tasks

Jiqun Liu
School of Communication & Information
Rutgers University
New Brunswick, NJ, USA, 08901
jl2033@scarletmail.rutgers.edu

Shawon Sarkar, Chirag Shah
Information School
University of Washington
Seattle, WA, USA, 98195
{ss288,chirags}@uw.edu

ABSTRACT

Complex search tasks that involve uncertain solution space and multi-round search iterations are integral to everyday life and information-intensive workplace practices, affecting how people learn, work, and resolve problematic situations. However, current search systems still face plenty of challenges when applied in supporting users engaging in complex search tasks. To address this issue, we seek to explore the dynamic nature of complex search tasks from process-oriented perspective by identifying and predicting implicit *task states*. Specifically, based upon the Web search logs and user annotation data (regarding information seeking intentions in local search steps, in-situ search problems, and help needed) collected from 132 search sessions in two controlled lab studies, we developed two task state frameworks based on intention state and problem-help state respectively and examined the connection between task states and search behaviors. We report that (1) complex search tasks of different types can be deconstructed and disambiguated based on the associated nonlinear state transition patterns; and (2) the identified task states that cover multiple subtle factors of user cognition can be predicted from search behavioral signals using supervised learning algorithms. This study reveals the way in which complex search tasks are unfolded and manifested in users' search interactions and paves the way for developing state-aware adaptive search supports and system evaluation frameworks.

CCS CONCEPTS

• **Information systems → Users and interactive retrieval**.

KEYWORDS

Complex search task; task state; interactive IR

ACM Reference Format:

Jiqun Liu and Shawon Sarkar, Chirag Shah. 2020. Identifying and Predicting the States of Complex Search Tasks. In *2020 Conference on Human Information Interaction and Retrieval (CHIIR'20), March 14–18, 2020, Vancouver, BC, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3343413.3377976

CHIIR '20, March 14–18, 2020, Vancouver, BC, Canada

ACM ISBN 978-1-4503-6892-6/20/03…$15.00
https://doi.org/10.1145/3343413.3377976

1 INTRODUCTION

Search systems are a major component of the intelligent assistance that is situated in broader sociotechnical ecosystems. People's interactions with search systems are often motivated by tasks that emerge from evolving, continuous problematic situations [5, 6]. Search systems and technologies have experienced phenomenal success in recent years, especially in addressing fact-finding and navigational search tasks [58]. However, current search systems, task and interaction models, and the underlying algorithms still face plenty of challenges when applied in supporting *complex tasks* which are intellectually challenging and involve multi-round, multidimensional search interactions (e.g., planning a research project, evaluating dental plans) [1, 18]. To address this problem, it is critical for interactive information retrieval (IIR) researchers to define an analytical, dynamic approach that is both theoretically meaningful and practically applicable to the anatomy of complex tasks.

The idea of classifying and conceptually deconstructing tasks is not new in information seeking and IR communities. Many existing task models (e.g., [13, 31, 34]) have jointly supported a large body of IIR studies concerning identifying and responding to users' tasks as static, overarching goals that motivates search actions. However, very little research has attempted to explore how search tasks are unfolded and evolve during the process of search interaction, and how we can optimize search system supports according to the dynamic states of complex search tasks.

Task as a multi-level concept in information seeking and IIR research can be defined with a *nested model* where search task is a subset of the associated information seeking task within the context of an overarching work task [11]. From a process-oriented perspective, the sequence and transitions of states in a search session often reveal essential properties of a search task, the associated work task as well as the task doer. In the light of Newell and Simon's human problem-solving framework [45], the transitions of states and actions can be considered as representations of users' iterative explorations in the evolving solution space behind the task.

Integrating the human problem-solving perspective with the static definitions of task complexity, we define *complex search task* as *search tasks that involve potentially broad, uncertain solution space or space of methods.* This uncertain solution space usually leaves limited potential for planned actions and is constantly shaped by multiple factors during a search session, such as: 1) predefined search goal(s); 2) users' search skills, topic and procedure knowledge regarding the task at hand, and the internal structure of information processing systems (IPS); 3) unanticipated results and search problems; 3) supports from the available search system(s).

In contrast to the static view of task complexity (e.g. [13, 31]), the process-oriented definition of complex search task is built upon a fundamental assumption: the complexity of a search task is not only determined by a set of predefined problems or desired goals (which may change during information seeking episodes), but also shaped by *users* and *the search systems they interact with*. In other words, we cannot fully reveal the nature of a complex search task without understanding the process of doing it.

Hence, from the process-oriented perspective, it is critical to investigate how people perform complex search tasks and explore the uncertain solution space behind the associated problem on multiple dimensions. At the operationalization level, we can dig deeper into the nature of complex search tasks through studying how these tasks are translated into and reflected in the patterns of intention states [38], cognitive biases [57], behavioral states in search session segments [25], encountered problems [56], level of search satisfaction [29] and other aspects of search interactions. In order to explore this in greater depth, this study aims to:

- Characterizing the dynamic nature of complex search tasks from a process-oriented, state-based perspective.
- Predicting the implicit task states from observable behavioral signals using supervised learning algorithms.

Based upon the empirical evidences from two user studies, we seek to address the above two research problems and advance knowledge in two aspects: 1) representing and explaining the dynamic aspect of complex search tasks using the distributions and transition patterns of task states; 2) revealing the connection between task states and users' search tactics and thereby paving the way towards building state-aware adaptive search supports.

2 RELATED WORK

2.1 Complex Search Tasks

An important branch of IIR research involves understanding and measuring the impacts of task facets on search behaviors, experiences, and performances [34]. *Task complexity* is one facet that has received considerable attention. Based on different task properties, researchers have developed multiple frameworks to define task complexity in the context of information seeking and searching.

For instance, Byström and Järvelin [12] studied the impacts of work task complexity on information seeking and use and developed a five-class complexity framework. Based on a qualitative investigation, they found that in complex tasks, the intentions of understanding, sense-making and problem formulation are essential and requires different types of information through a variety of information sources at different points of information seeking epsides. Kelly et al. [31] explored the cognitive complexity of task and adopted Bloom [8]'s taxonomy of learning domains in characterizing different levels of task complexity. Their results indicate that complex search tasks required significantly more search activities from users. Urgo et al. [54] further extended Kelly's work by integrating knowledge dimension with learning dimension in task design and learning assessments. Capra and his colleagues used task prior determinability (i.e. the level of uncertainty about task outcomes and processes) as a representation of task complexity and argued that the variations in needed items and the clarity of dimensions for result evaluation significantly affects task determinability [13, 14]. Liu et al. [38] extracted two major static task facets, *task product* and *task goal*, from Li and Belkin [34]'s framework and used the combination of these two facets to define task complexity. They found that tasks of different levels of complexity can be represented by different patterns of local intentions. Similarly, He and Yilmaz [24] also employed multiple task facets to identify and disambiguate tasks of different types in field settings.

The last few years have seen the interactive IR communities tackle more complex search tasks that involve multiple rounds of search iterations [18, 36]. Many of the existing studies represented different levels of task complexity using one or more static task features and revealed some of its behavioral effects (e.g., [17, 31, 37]). However, there has been little data-driven work representing complex search tasks as sequences of cognitive states in forms suitable for computational modeling. As a result, we still lack an effective approach to exploring the connections between predefined, static task properties and the dynamic transitions of task states.

2.2 Process Models of Search Tasks

When conceptualizing tasks from process-oriented perspective, we are essentially focusing on the process of *doing tasks*. In the IR community, a number of models and techniques have been developed to describe and explain different aspects of tasks and search activities. Bates [2] proposed the berrypicking model and argued that single-query, best-match model cannot capture the evolving nature of search tasks. In contrast to the traditional single-query model of ad hoc IR, berrypicking model illustrates the interactive process of searching and has been empirically supported by many task-based information seeking studies [20, 32, 47, 52]. Based on Kuhlthau's ISP model as well as a series of empirical works, Vakkari [55] developed a general framework of task-based information searching which consists of three stages: *pre-focus*, *focus formulation*, and *post-focus*. His studies also indicate that there is a close association between the participants' problem states in task performance and the information need, the search tactics employed and the assessment of document relevance and utility. Similarly, Belkin [4] proposed a conceptual model that represents session-level information seeking episode as a sequence of users' iterative interactions with an interactive search system and the retrieved information objects.

The classical models discussed above are widely applied in describing the process and stages of tasks, information seeking and searching. However, Most, if not all of them, offers limited implications for building computational frameworks of task processes and developing dynamic supports for complex tasks at different moments [35]. To address this issue, some researchers seek to develop computationally-congenial models for representing task states and simulating task-based search interactions. For instance, Cole et al. [17] investigated user activity patterns in tasks of different types and demonstrated that task types and levels of task difficulty can be disambiguated by the sequences of user activity states (derived from page visiting behaviors) and cognitive processing states (approximated using eye movement patterns). To develop an effective formal model of search interactions, Fuhr [23] proposed the IIR-PRP for extending probabilistic IR to IIR context and representing users' situation transitions at different moments of information searching episodes. The IIR-PRP model serves as an important step towards building a computational framework for supporting the functional design of search systems. However, this model abstracts out a variety of user characteristics and lacks effective representations of the task states and associated cognitive variations.

2.3 Intentions, Problems, and Help Needed

Users often seek to accomplish different things (e.g., find specific item, explore a new topic) at different points of task-based search interactions. These *local goals* that users try to achieve in individual query segments or search iterations within the context of a global search task can be defined as *information seeking intentions* [38, 43]. In complex search tasks that involve multiple rounds of search iterations, the combinations of information seeking intentions can be considered as a representation of the *active, planned* dimension of task states. In this sense, the changes in intention combinations can partially reveal the implicit transitions of task states.

The IR community's exploration of information seeking intentions started with theoretical research on classifying short search sessions. For instance, Broder [10] developed a search session typology which consists of three categories: navigational, transactional, and informational. Similarly, Kellar, Watters, and Shepherd [30] proposed a classification scheme of users' intentions, including fact findings, information gathering, browsing, and transactions. Beyond the theoretical speculation on intention classification, Xie [59] empirically studied users' motivations in information seeking and search interactions and identified a set of interactive intentions and search tactics. In the context of Web search, Jansen and Booth [28] developed a three-level hierarchy of user intent, aiming to automatically classify Web search queries based upon the intentions behind queries. Rha et al. [46] proposed a typology of twenty information seeking intentions based on Xie [59]'s work and explored the connection between users' intentions and their associated query reformulation strategies. To develop a more comprehensive understanding of Web search activities, Liu et al. [38, 40] investigated the connections between static search task features (i.e. task product and task goal), the distribution and transitions of information seeking intentions, and users' search behaviors and performance.

In addition the the active dimension of search tasks, to gain a deeper understanding of search interactions, many IIR researchers have also explored the *unanticipated, situational* aspect which mainly covers users' *search problems* and *help needed* for addressing the problems in tasks of different types [22, 50]. The common barriers that users face when searching for information include internal barriers (e.g., lack of knowledge, unable to articulate information need) [7], external barriers (e.g., time constraints, institutional restrictions) [49], interpersonal barriers (e.g., lack of help from other people) [53], and other types of barriers [16]. Previous research also identified various possible causes that lead to these barriers, such as users' lack of domain and topic knowledge, system knowledge, and necessary search skills. Regarding help needed, when people come across any problem while looking for information, they often seek help or supports of different types from system and/or people [27]. Existing studies showed that help-seeking situations are primarily influenced by users' personal and cognitive characteristics as well as the nature of task [60]. Based on the existing work on search barriers and help-seeking behaviors, Sarkar et al. [48] investigated the connection between the situational aspect of task and users' behaviors and demonstrated that search behavioral signals can inform us about users' in-situ search problems and supports preferred at different points of task process.

In sum, to develop adaptive supports for complex search tasks, it is critical to construct a *task state framework* which speaks to the implicit cognitive variations in task process and can support the computational modeling of state transitions and search interactions. Previous explorations on users' information seeking intentions, in-situ problems and help needed shed lights on the dynamic nature of complex search tasks. From process-oriented perspective, the factors from both active and unanticipated dimensions can be considered as basic ingredients of task states. Investigating the combinations of these factors and the transitions of the combinations can enhance our understanding of how complex search tasks are unfolded in users' explorations of the uncertain solution space.

3 RESEARCH QUESTIONS

To address the research gaps discussed above and develop a state-based framework of complex search tasks, this work reports on two controlled-lab-based studies which investigated the active dimension and situational, unanticipated dimension of task states respectively. These two studies collected data on users' local information seeking intentions, encountered problems and help needed in complex search tasks of different types. Our work sought to answer following research questions (RQs):

- **RQ1**: What are the states of complex search tasks?
- **RQ2**: What are the transition probabilities between the states in complex search tasks of different types?
- **RQ3**: To what extent can we predict search task states from Web search behavior?

Among the RQs above, RQ1 and RQ2 focus on identifying and extracting dynamic task states from intentions, encountered problems and help needed at different points of search and characterizing the state transition patterns in varying task types. RQ3 speaks to the connection between behavioral signals and implicit task states and seeks to inform state detection and adaptive support development.

4 METHODOLOGY

This work includes two controlled lab studies, *information seeking intention* (ISI) study and *problem-help* (PH) study. ISI study explored users' information seeking intentions and search actions in different query segments of complex search tasks, and PH study investigated the association between users' encountered problems, help needed, as well as their search behavior. Analyzing the empirical evidences collected from these two studies enabled us to characterize and model the states of complex search tasks from different perspectives.

4.1 Information Seeking Intention Study

4.1.1 Information Seeking Intentions. To collect data on users' information seeking intentions in search iterations, this study employed the search intention typology developed by Rha et al. [46] based on a subset of Xie [59]'s classification scheme of interactive intentions. The researchers gave a detailed description of this typology to the participants before their search sessions were replayed for intention annotation. Then, participants were asked to identify their intention(s) for *each query segment* (all that occurred between one query and the next, including the queries) based on the typology. Participants could identify multiple intentions for the query segments where they sought to achieve multiple local goals. Thus, the intentions identified by users were not mutually exclusive and should be considered as different dimensions of task states. Table 1 presents the typology of twenty information seeking intentions.

4.1.2 Complex Search Tasks and Participants. To simulate complex search task context and elicit rich search interactions and intention transitions, we designed four tasks within the domain of journalism, including copy editing (CPE), story pitch (STP), relationships (REL), and interview preparation (INT). To minimize the potential effect of knowledge background, we recruited forty undergraduate students majoring in journalism from a research university.

The four types of complex search tasks were designed mainly based on two facets from Li and Belkin [34]'s classification scheme: *Product* and *Goal*. Regarding task product, *intellectual task* refers to a task which motivates people to produce new ideas or plans, whereas *factual task* mainly focuses on locating objective facts or information items. In terms of goal, task with *specific goal* refers to a task with a goal that is explicit and measurable. Task with *amorphous goal* has no explicitly defined process or outcome. Each task type has two versions corresponding to two different topics (i.e. coelacanths; methane clathrates and global warming) that our participants were thought likely to not be familiar with. We chose these two topics to control the potential impact of topic familiarity and also to better simulate the context of complex search tasks. Each of the participants was asked to conduct two search tasks of varying topics and types, in Latin Square design, pairing CPE with INT, and STP with REL, to balance tasks by topic and facet values.

The four search tasks (Coelacanth topic) are presented as follows:

- **Copy Editing** (Factual Specific): Assignment: You are a copy editor at a newspaper and you have only 20 minutes to check the accuracy of six italicized statements in the excerpt of a piece of news story below. Task: Please find and save an authoritative page that either confirms or disconfirms each statement.
- **Story Pitch** (Factual Amorphous): Assignment: You are planning to pitch a science story to your editor and need to identify interesting facts about the coelacanth ("see-la-kanth"), a fish that dates from the time of dinosaurs and was thought to be extinct. Task: Find and save Web pages that contain the six most interesting facts about coelacanths and/or research about their preservation.
- **Relationship** (Intellectual Amorphous): You are writing an article about coelacanths and conservation efforts. You have found an interesting article about coelacanths but you need to be able to explain the relationship between key facts you have learned. Task: In the following, there are five italicized passages, find an authoritative Web page that explains the relationship between two of the italicized facts.
- **Interview Prepration** (Intellectual Amorphous): You are writing an article that profiles a scientist and their research work. You are preparing to interview Mark Erdmann, a marine biologist, about coelacanths and conservation programs. Task: Identify and save authoritative Web pages for the following: Identify two (living) people who likely can provide some personal stories about Dr. Erdmann and his work. Find the three most interesting facts about Dr. Erdmann's research. Find an interesting impact of Dr. Erdmann's work.

4.1.3 Study Procedure. The user study started with a demographic questionnaire and a tutorial video on the additional interface features offered by our browser plugin. Participants were free to search anywhere on the web, with the only restriction being to conduct their searches in the browser with our plug-in turned on for logging their search actions and enabling them to save pages. After reading the task description and answering questions about their topic- and task-related knowledge, participants had up to 20 minutes to complete the first assigned task by searching on the web and bookmarking useful pages. Participants could choose to enter the next phase early if they completed it to the best of their ability.

Afterwards, we asked participants to complete a post-task questionnaire on their overall search experience. Then, they were asked to read a guidance of the intention annotation task and to watch a short video explaining how to annotate information seeking intentions. We then replayed the entire search session, query segment by query segment, asking for intention annotation for *every segment*, in sequence. The intention annotation assignment had no time limit. Participants were asked to choose which intentions applied to each query segment in the search session based on the predefined typology. Within each query segment, participants could select any number of intentions from the typology (see Table 1). Participants could choose "other" if their intention did not match the 20 intentions listed. This intention selection process was repeated for every query segment during the process of intention annotation. The same study procedure was then followed for the second task, and the study session ended with an exit interview with open-ended questions related to participants' search experience and performance. The entire user study process took about two hours for each participant. More detailed descriptions regarding the questionnaires and intention annotation interface are reported in [40, 46].

4.2 Problem-Help Study

4.2.1 In-situ Search Problems and Help Needed. To gather empirical evidences concerning users' search problems and help preferred at different points of search, this study developed a list of problems and possible supports based on the findings from previous IIR studies on search barriers and help-seeking activities (e.g., [16, 50, 60]). We introduced the problem-help list/typology to participants and asked them to identify the problem they encountered and help needed based on our list every time before they intended to formulate a query during their search sessions. Similar to intention annotation in the ISI study, participants in the PH study could identify multiple types of problems and help for applicable query segments. Table 2 presents the list of questions designed for the PH study.

4.2.2 Complex Search Tasks and Participants. Similar to ISI study, in the PH study we designed two search tasks by manipulating the type of information need (cognitive or social), task goal, and task product based on Li and Belkin [34]'s classification framework. In addition, we asked every participant to complete a 5-min warm-up task before they started working on the two "formal", 20-min complex search tasks, aiming to familiarize them with the study procedure (especially the in-situ problem-help annotation and useful document identification), search interface and lab environment. Adding the warm-up task is critical for gathering reliable data here as the in-situ annotation is not part of the natural search process and participants might need extra time to adapt to the change [39].

The description of search tasks are presented as follows:

Table 1: Typology of Information seeking intentions.

Type of Intention	Information Seeking Intentions
Keep record	Keep record of a link (KR)
Identify search information	Identifying something to start (IS); Identify something more to search (IM)
Learn	Learn domain knowledge (LK); Learn database content (LD)
Find	Find known item(s) (FK); Find specific information (FS); Find items sharing a named feature (FN); Find items without predefined criteria (FW)
Access item(s)	Access a specific item (AS); Access items with common characteristics (AC); Access a website/homepage or similar (AW)
Evaluate	Evaluate correctness of an item (EC); Evaluate usefulness of an item (EU); Pick best items from all the useful ones (EB); Evaluate specificity of an item (ES); Evaluate duplication of an item (ED) (i.e., determine whether the information in one item is the same as in others)
Obtain	Obtain specific information to highlight or copy (OS); Obtain part of an item (OP); Obtain a whole item(s) (OW)

Table 2: The question for eliciting potential in-situ search problems and help needed, and the possible responses.

Question:*What problems are you facing at this moment? Select all that apply.*
- 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 enough about the topic.
- I am feeling impatient.
- 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.
- There is just too much information.
- What I am looking for does not seem to be available.
- No problem encountered.
Question:*What kind of things would help you at this moment? Select all that apply.*
- Recommendations by the system about useful search queries.
- Recommendations by the system about potentially useful web pages.
- Recommendations about useful search steps and strategies.
- Find me people (e.g., domain expert) who may be able to help.
- 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.

- **5-min Warm-up Task** (Factual Amorphous): 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?
- **Cognitive Task** (Intellectual Amorphous): 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.
- **Social Task** (Intellectual Amorphous): You will be attending a 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 based on your estimation of the other guests' interests, preferences and backgrounds. To be fully prepared, please create a list of at least five interesting up-to-date topics. For each topic, please identify the useful web pages and a very brief explanation for why you choose this as a potentially suitable topic.

4.2.3 Study Procedure. The PH user study started with a brief questionnaire provided by the customized browser plugin-in about participants' task and topic familiarity as well as a series of demographic factors. Then, participants had up to 20 minutes to complete each assigned search task. During the search session, when a participant was about to formulate a new query during the search process, a pop-up window appeared with the problem-help questionnaire (see Table 2), asking the participant to report his or her in-situ search problem(s) at the moment and the help needed for resolving the problems. After each search task, participants were asked to complete a short post-search questionnaire about their perceived task difficulty and overall search experience. Participants' search activities (e.g., query, timestamp, clicks, URLs, page type) were recorded with a Google chrome browser plugin and Morae. Simiar to the ISI study, the PH study session ended with an exit interview where we asked open-ended questions about participants' encountered problems and why they prefer certain types of help at different points of search sessions. The entire user study process took about one and a half hours for each participant. More detailed descriptions regarding the task-related questionnaires and problem-help annotation interface are offered in [48].

4.3 Measurements and Data Analysis

4.3.1 RQ1: Identifying Task States. The data collected from ISI and PH studies enabled us to model the states of complex search tasks from active aspect and situational, unanticipated aspect respectively. In the datasets, each intention, problem, and help item was represented using a unique binary variable (present=1, absent=0). For state identification, we used *K-modes clustering analysis* for extracting clusters out of user annotation data. K-modes clustering as a unsupervised learning method extends the traditional K-means paradigm to cluster categorical data [15]. In the clustering analysis, different information seeking intentions, in-situ problems, and types of help needed were considered as separate elements within the vectors representing unique task states.

To test the validity of the task state categories extracted from annotation data, we ran *external judgment of state types* with two external assessors. Specifically, we randomly extracted 10% of searches from each type of tasks and ask the two assessors to manually annotate task state for each query segment independently given the four-state typology we defined. Each assessor were provided with the video of participants' search process, the intention or problem-help annotation and search behavior data, and the clustering results (i.e. the specific task state a given query belongs to) were removed. To measure the validity of task state labels, we computed three *Cohen's Kappa coefficients*, between 1) the two annotators, 2) the annotator *A* and the clustering algorithm, and 3) the annotator *B* and the clustering algorithm. To ensure the quality of task state labeling and judgment, we recruited two advanced Ph.D students majoring in IR as our external state annotators here.

4.3.2 RQ2: Understanding State Transition Patterns. After we identified and validated task state labels, we computed the transition probabilities between different task states. Focusing on the process-oriented, dynamic aspect of tasks, we aim to reveal the nature of complex search tasks of varying types by investigating the difference in their task state transition patterns.

4.3.3 RQ3: Predicting Task States from Search Behaviors. To inform the development of state-aware adaptive search supports, RQ3 seeks to examine the extent to which we can predict implicit task states from behavioral signals. To answer this question, we employed multiple supervised learning algorithms and utilized a series of widely-used search behavioral measures for building classifiers.

The search behavioral features include: *query behavior*: query length, query reformulation type; *browsing behavior*: number of clicks; number of content pages visited; *dwell time (second)*: mean dwell time on each search engine result page (SERP); mean dwell time on each content page; total dwell time on content pages; and *usefulness judgment*: number of bookmarks. Based on these basic measures, we adopted three types of feature sets for predicting task states: 1) behavioral measures in *current query segment* associated with the target task state, 2) *session-level behavioral measures* before current query segment, and 3) the *combination of 1) and 2) sets*. Regarding the session-level behavioral measures, we computed the average values of behavioral measures for the associated search session (before current query segment).

5 RESULTS

To answer the proposed three RQs, we analyzed the search behavioral data and user annotations collected from ISI and PH studies. In ISI study we collected data from 693 query segments generated by 40 participants in 80 task-based search sessions. The PH study elicited data from 273 query segments generated by 26 participants. The search sessions (measured by number of query segments) are relatively long (mean length in ISI study: 8.66; mean length in PH study: 5.25), indicating that the simulated complex search tasks were successful in eliciting rich search iterations. To clarify the contribution and implication of our study, in this section, we organize the results from data analyses according to the proposed RQs.

5.1 RQ1: Identify Task States

The identification of task states started with K-modes clustering analysis. Before that, we employed average silhouette method to determine the optimal number of clusters. We extracted four clusters as separate task states from the ISI dataset and six clusters from the PH dataset. The clustering analysis for PH study was conducted based on 216 query segments as the problem-help annotation was missing for some of the repeated queries due to system errors.

focusing on the active, intention aspect of task state, we identified the following four states of complex search tasks. We interpreted each extracted task state based on the main (most frequent) information seeking intentions within the state.

- **Exploitation (E)** (frequency: 54.3%, 376 query segments): The two most frequent intentions are *find specific information* (39.4%) and *identify something more to search* (40.4%). Meanwhile, the intention of identifying something to start searching never occurs. In this state, users may have a clear topic in mind and they try to follow the current search path, keep exploiting the information patch at hand and search for more relevant pages.
- **Known-Item (K)** (frequency: 18.2%, 126 query segments): The two most frequent intentions are *find specific information* (100%) and *obtain specific information items* (100%). In this state, users may have very specific, well-defined information need(s) or item(s) in mind.
- **Exploratory (EX)** (frequency: 16.6%, 115 query segments): The most frequent intention in this state is identify something to start searching (100%). In this state, users may try to adopt new search strategies, explore unknown subtopics, or open new search paths.
- **Learn and Evaluate (L)** (frequency: 10.9%, 76 query segments): In this state, most intentions under the *Evaluate* category (above 60%) and the intentions of learning domain knowledge and keeping useful links (both above 80%) occurred frequently.

Similarly, with respect to the situational (problem-help) aspect of task state, we identified six task states and explained them based upon the most frequent search problem(s) and/or help needed. We use acronym to represent each state here as it is difficult to assign any meaningful label to cover all traits of these problem-help states.

- **IO-P** (frequency: 21.3%, 46 query segments): The most frequently occurred problem is information overload (IO) (34.8%) and main type of help needed is Web page (P) recommendation (74%).
- **ASK-LT-PE** (frequency: 11.6%, 25 query segments): In this state, users are very likely to experience the anomalous state of knowledge [3] (ASK: do not know how to express their information need or what exactly they are looking for) (64%) and other barriers, such as lack of topic knowledge (LT) (72%) and not knowing potentially useful information sources (64%). In this state, they usually prefer to have people (PE) who can guide them through the search process.
- **ASK-SU-M** (frequency: 11.6%, 25 query segments): In this state, users are very likely to encounter the ASK issue (76%) and the problem of not knowing useful sources (80%). Here, users often prefer to have multiple types of supports, such as page recommendation (88%), query recommendation (96%), and strategy recommendation (92%).
- **NP** (frequency: 36.1%, 78 query segments): In this state, users often have no explicit search problem (NP) (70%) and thus do not need any specific help from the search system (88.6%).
- **LT-M** (frequency: 4.6%, 10 query segments): In this state, the problem of lacking topic knowledge frequency occurs (89%) and users need multiple types of help, such as page recommendation (89%), people recommendation (89%), and search strategy recommendation (100%).
- **SU-QU** (frequency: 14.8%, 32 query segments): In this state, users are very likely to encounter the problem of not knowing useful information sources (63%) and usually prefer to have useful query recommendations from the system (75%).

Table 3: Behavioral variations across different intention-based task states: (*: p<.05, **: p<.01).

behavior	Dunn's posthoc test
querylength*	E>EX*,K>EX*,E>L*,K>L*
dwellSERP**	K>EX*,E>L*,K>L**
dwcontent**	K>E*,K>EX*,L>E*,L>EX*
N.content**	E>L*,K>L*, E>EX**, K>EX**
totalcontent**	K>E*,K>EX*,L>E**,L>EX*
N.clicks**	L>E**,L>K**,L>EX*, EX>E*
N.bookmark**	L>E**, L>EX**, L>K*

To test the validity of the above task states extracted by K-modes clustering algorithm, we invited two assessors to do manual task state annotation and computed the Cohen's Kappa coefficients κ for all three pairs: 1) annotator A and annotator B: 0.782 (ISI-based state), 0.768 (PH-based state); 2) annotator A and clustering algorithm: 0.716 (ISI-based state), 0.717 (PH-based state); 3) annotator B and clustering algorithm: 0.744 (ISI-based state), 0.682 (PH-based state). The Cohen's Kappa agreements in all pairs are above 0.65, which is considered *substantial* agreement [33]. This high level of agreement demonstrates that the task state typology generated by the clustering algorithm is reliable and can be used for further analysis. Also, it is worth noting that neither of the between-annotator agreements crosses the threshold of "almost perfect" agreement (0.8) [33], indicating that inferring implicit task states from search interactions is not an easy job (even for human annotators).

To further explore the boundaries between task states, we examined the extent to which the identified states differ from each other in terms of the associated search behaviors. We used *Kruskall-Wallis test* to test the between-state behavioral differences as the results of *Shapiro-Wilk tests* indicated that none of the search behavioral data was normally distributed. Since we have multiple groups (states) identified in both ISI and PH studies, we employed *Dunn's test* with Benjamini-Hochberg correction for post hoc pairwise analyses.

Table 3 presents the results of Kruskal-Wallis and Dunn's post hoc tests on the behavioral variation across different intention-based task states. The descriptive statistics (median and IQR) are omitted here for brevity. In general, when participants had a relatively clear topic or specific item in mind (in exploitation or known-item states), they tended to issue longer, more specific queries and spend more time on seeking for most relevant information directly on SERPs. In contrast, when participants were in learning and evaluation state, they tended to stay longer on content pages and do more clicks and bookmarks (for usefulness judgments and result evaluation). These results demonstrate that the transitions in intention-based task states are closely associated with the variations in participants' search tactics in local search steps.

Table 4: Behavioral variations across different problem-help-based task states (*: p<.05, **: p<.01).

behavior	Dunn's posthoc test
querylength*	ASK-SU-M>IO-P**, ASK-LT-PE*, NP*, SU-QU*
dwellSERP**	ASK-LT-PE>IO-P**, NP**,LT-M*, SU-OU*; ASK-SU-M>NP*, SU-OU**
dwcontent	N.A.
N.content*	ASK-LT-PE>IO-P*, ASK-SU-M*, NP*, SU-QU*; ASK-SU-M>NP*
totalcontent**	ASK-LT-PE>IO-P*, ASK-SU-M*, NP*, LT-M*, SU-QU*
N.clicks**	IO-P>ASK-SU-M*; ASK-LT-PE>ASK-SU-M*, LT-M*, SU-QU*
N.bookmark**	NP>ASK-LT-PE**, ASK-SU-M*, LT-M*, SU-QU*

Table 4 illustrates the behavioral variations across different problem-help states. The descriptive statistics (median and IQR) are omitted here for brevity. The results indicate that we participants encountered the problems of ASK and lacking topic knowledge (ASK-LT-PE), they tended to be more active in browsing SERPs and reading content pages, seeking to find useful cues for formulating queries and deciding right search paths. When participants encountered the information overload problem (IO-P), they were likely to be distracted by many (irrelevant) information items, which resulted in more clicking actions. In contrast, when participants had no explicit search problem, they tended to bookmark more useful pages, indicating that they were on the right track of searching.

The above analysis on helps clarify the boundaries between different task states at behavioral level and thereby paves the way for predicting task states from search behavioral signals (RQ3).

5.2 RQ2: Understand State Transition Patterns

Aiming to go beyond predefined task properties (e.g. task facets) and explore the dynamic aspect of complex search tasks, we examined the state transition patterns in tasks of different types. From process-oriented perspective, the difference in state transition pattern represents the divergence in the way in which people explore the uncertain solution space associated with the task and thus may help disambiguate different types of complex search tasks. Modeling state transition patterns can enhance our understanding of how predefined task facets and the combination of facets are manifested dynamically in search sessions. Figures 1-4 illustrates the state transition patterns of four types of complex search tasks assigned in the ISI study, and Figures 5-6 presents the state transition patterns of the two task types in PH study.

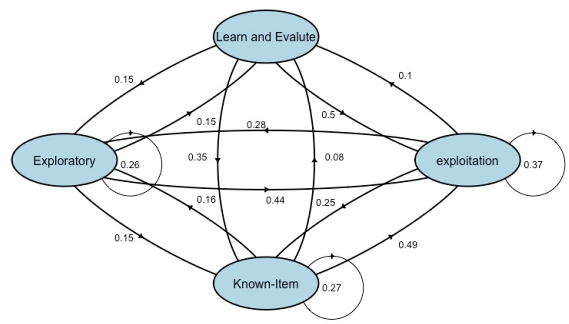

Figure 1: ISI copy editing task (factual specific).

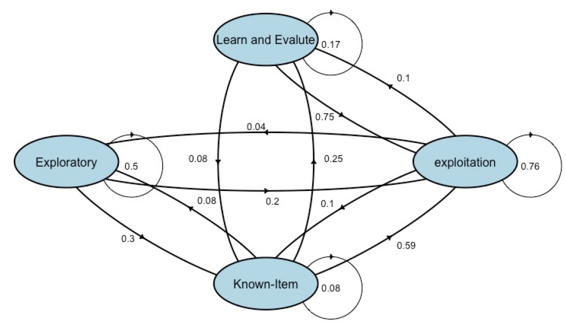

Figure 2: ISI story pitch task (factual amorphous).

Overall, our results demonstrate that the process of doing a complex search task is usually *nonlinear*. In all six tasks, we observed transition loops both between and within task states (i.e. remaining in the same state). The difference in task type (defined by the combination of task facets, cf. [34]) was also reflected in the variations of task state transition probabilities. For instance, compared to the copy editing task (factual specific), the story pitch task with amorphous task goal motivated participants to do more exploratory, open-ended search (i.e. 50% chance of remaining in the exploratory state). In copy editing task, participants searched for known information items more frequently but rarely stayed in the learn and evaluate state. Also, participants working on copy editing task transited from learn and exploitation states to known-item search more frequently. This finding indicates that in factual specific task, it might be easier for searchers to identify and extract specific information items from previous learning, evaluation, and topic exploitation process. In story pitch task, participants often stayed within exploitation state and kept exploring the topic at hand due to the ambiguity of task goal.

In the two intellectual amorphous tasks, participants tended to transit more actively between exploratory state and learn and evaluate state, and remained in these two states more frequently. Note that in the story pitch task, participants never transit between exploratory and learning states. Similarly, in copy editing task, participants never stayed in the learn and evaluate state for two continuous search iterations. This may be because the two intellectual tasks motivated participants to take more information-literate actions (e.g., learning connections between facts, evaluating usefulness of pages, exploring new information cues) for producing the intellectual product(s). Besides, in the three goal-amorphous tasks, participants remained in the exploitation state more frequently (rather than frequently transit to known-item state) due to the difficulty of searching with a vague, ill-defined goal.

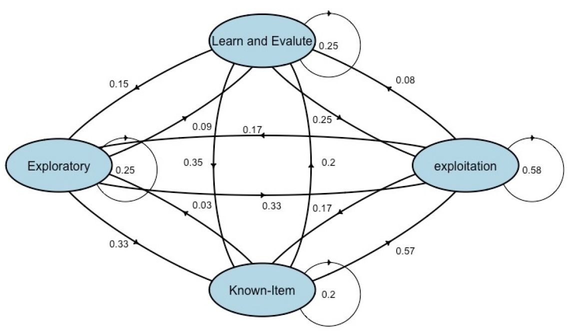

Figure 3: ISI relationship task (intellectual amorphous).

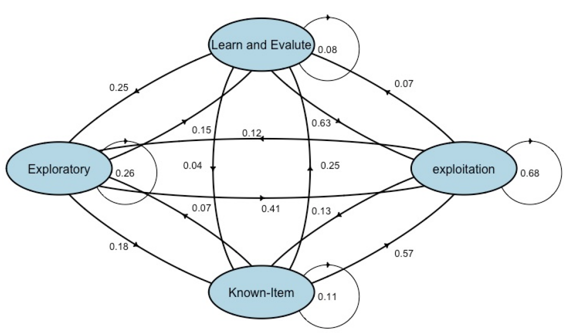

Figure 4: ISI interview prep task (intellectual amorphous).

The problem-help-based task states offers us a different perspective on the dynamic process of complex search tasks. Figures 5 and 6 illustrate problem-help state transition patterns. Since the problem-help approach produced six separate task states, to improve the clarity and readability of these two figures, we omitted the edges/transitions with a probability lower than 20%.

Overall, in the cognitive task, participants remained in the two ASK-related states more frequently (ASK-SU-M: 40%, ASK-LT-PE: 44%), indicating that expressing information need(s) with query terms is a major challenge here. Also, instead of simply transiting to the NP (no clear problem encountered, no help needed) state, participants frequently moved from the two ASK states to other problematic states, such as SU-QU (unaware of useful sources) and IO-P (information overload). In particular, the ASK-LT-PE state did not even have a direct transition path toward NP state (i.e. no edge between the two states). These results demonstrate that participants were not well supported in this complex search task. Going back to the original search session videos, we found that many participants started with copying part of the task description as search queries (hoping to get direct answers to the question) but unfortunately received bad (irrelevant) results. After that, they tried to formulate queries based on their own understanding of the task and still encountered plenty of barriers. This might be because this cognitive task required participants to build a bridge between two topics from completely different domains (i.e. Miss Universe competition and Indian government's policies) , which is intellectually challenging in the context of Web search.

In contrast, the social task appeared to be less complicated as the probability of transition from ASK-LT-PE to NP was 44%, which is much higher than that of the cognitive task (20%). Another possible evidence of relatively low complexity is that participants remained in the NP state more frequently in the social task (69%). In addition, the ASK-SU-M and LT-M states never occurred in the social task, indicating that participants were better supported by the system in searching for information to satisfy their social needs.

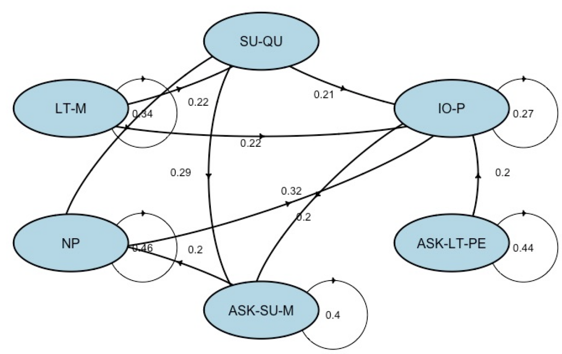

Figure 5: PH cognitive task (intellectual amorphous).

5.3 RQ3: Predict Task States from Behaviors

The state transition patterns discussed above shed light on the dynamic nature of complex search tasks. To develop state-aware adaptive support for users, it is critical to examine the extent to which we can predict task states from behavioral signals. To answer RQ3, we built several classifiers based upon the behavioral features introduced in the section 4.3.3. We trained and evaluated classifiers with an 80/20 split on training/testing data and compared them with two baselines: 1) random baseline and 2) most frequent labeling.

The findings presented in Table 5 and Table 6 show that: 1) overall, the best performers/classifiers built on behavioral features significantly outperform the corresponding baseline models in the overall accuracy of predicting task states; 2) constructing classifiers

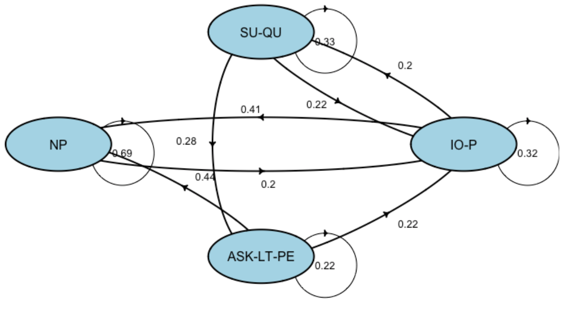

Figure 6: PH social task (intellectual amorphous).

Table 5: Accuracy score of task state prediction (ISI).

Classifier	Current seg.	Prev. session	All data
Logistic Regression	0.588**	0.583**	**0.594****
Support Vector Machine	0.527	0.535	0.547
XGBoost	0.559	0.535	0.562**
Random Forest	0.539	0.530	0.541
Decision Tree	0.433	0.410	0.427
Most Frequent	0.543	0.543	0.543
Random	0.249	0.249	0.249

Note: Significant values indicate whether the predictor is significantly better than the best baseline (*:p<.05, **:p<.01). The best performer is boldfaced.

using behavioral data from current query segment only can always outperform the baseline models; 3) using previous-session-based classifiers, it is possible to predict task states with an accuracy score significantly higher than that of the best baseline. Note that in problem-help state prediction, the classification task involves six different states. This typology contains distinctions that could be too fine for a future interactive system to disambiguate. Thus, based on the frequency distribution of specific features, we collapsed six specific P-H states into three types. We found that our models achieved better performances in this "lower resolution" prediction, with the best performer reaching almost 70% accuracy. As our response to RQ3, these results jointly illustrate the potential of search behavioral models in predicting dynamic task states and empirically proves that it is possible to monitor state transitions in search and to develop state-aware adaptive system supports.

Table 6: Accuracy score of task state prediction (PH).

Classifier	Current seg.	Prev. session	All data
Logistic Regression	0.511**	0.359	0.416**
Support Vector Machine	0.366	0.389	0.384
XGBoost	0.543**	0.522**	0.572**
Random Forest	0.580**	0.546**	0.574**
Decision Tree	**0.587****	0.552**	0.577**
Most Frequent	0.361	0.361	0.361
Random	0.166	0.166	0.166

6 DISCUSSION AND CONCLUSION

To explore the dynamic aspect of complex search tasks and understand how predefined task properties are manifested dynamically in search sessions, we conducted two controlled lab studies and constructed task state framework based on users' information seeking intentions, in-situ search problems and help needed in local steps. The finding of this study enhances our understanding of complex search tasks from process-oriented perspective and demonstrates that the abstract concept of task complexity can be embodied by the transitions and variations of task states and search tactics. Regarding the RQs, we have following answers.

To answer the RQ1 (identify task states) and RQ2 (understand state transition patterns), we extracted task states from annotation data using clustering algorithms and validated the cluster/state labels through manual annotations and assessments. The two state frameworks we developed cover both the active, intention dimension and the unanticipated, situational dimension of task states. Then, we examined the state transition probabilities in complex search tasks and demonstrated that the difference in task type can be detected from the variations in state transition patterns. These findings enrich our understanding of the nature of complex search task and extend the existing descriptive and computational models of task-based search process (e.g., [2, 23, 55]) by better revealing the subtle cognitive, situational changes in users' exploration of uncertain, evolving solution space and illustrating the *nonlinearity* (e.g., loops, state repetitions) of task completion process.

With respect to RQ3, results indicate that both intention-based and problem-help-based task states can be inferred and predicted using classifiers built on search behavioral features. Therefore, it is possible for intelligent search systems to detect task states in an online fashion and leverage the knowledge of task state in adaptively supporting searchers who have different local information seeking intentions and/or encounter search barriers of different kinds at different states of complex search tasks. In addition, the knowledge of task states can also be incorporated into user-centered system evaluation process and facilitate state-based search evaluation.

As always, there are limits to our work as well as needs for future efforts. This study only studied three elements of task states (i.e. intention, search problem, help needed) and left out other aspects that might significantly affect task completion process (e.g., emotional and knowledge states [42, 51]). However, this research speaks to a promising direction of conceptually deconstructing and computationally modeling complex search tasks from process-oriented perspective and may encourage future research to explore the changes and transitions on other dimensions and further enrich the task state framework. In addition, the findings reported here from two small scale user studies need to be tested based on the datasets collected from different complex task contexts (e.g. structured information search in specialist databases [9, 41]) and study settings (e.g., home environment [19]). It is also critical to investigate how user traits (e.g., search skills, cognitive limits) and other contextual factors (e.g. network latency) affect the distribution and transition of task states at multiple levels. Based on larger scale datasets and more fine-grained features (e.g., cursor movement signals [26], neuro-physiological features [21, 44]), it seems realistic to believe that we will eventually be able to develop adaptive search systems that can provide reactive and even proactive supports for users according to the prediction of task states.

7 ACKNOWLEDGEMENTS

This work is supported by the National Science Foundation (NSF) grant IIS-1717488. It also uses datasets generated as a part of the project supported by the NSF grant IIS-1423239.

REFERENCES

[1] A. H. Awadallah, C. Gurrin, M. Sanderson, and R. W. White. Task intelligence workshop @ WSDM. In *Proceedings of ACM International Conference on Web Search and Data Mining*. ACM, 2019.

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

[3] N. J. Belkin. Anomalous states of knowledge as a basis for information retrieval. *Canadian journal of information science*, 5(1):133–143, 1980.

[4] N. J. Belkin. Intelligent information retrieval: whose intelligence? *ISI*, 96:25–31, 1996.

[5] N. J. Belkin. Some (what) grand challenges for information retrieval. In *ACM SIGIR Forum*, volume 42, pages 47–54. ACM, 2008.

[6] N. J. Belkin. People, interacting with information. In *ACM SIGIR Forum*, volume 49, pages 13–27. ACM, 2016.

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

[8] B. S. Bloom et al. Taxonomy of educational objectives. vol. 1: Cognitive domain. *New York: McKay*, pages 20–24, 1956.

[9] D. Bonino, A. Ciaramella, and F. Corno. Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. *World Patent Information*, 32(1):30–38, 2010.

[10] A. Broder. A taxonomy of web search. In *ACM Sigir forum*, volume 36, pages 3–10. ACM, 2002.

[11] K. Byström and P. Hansen. Conceptual framework for tasks in information studies. *Journal of the American Society for Information science and Technology*, 56(10):1050–1061, 2005.

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

[13] R. Capra, J. Arguello, H. O'Brien, Y. Li, and B. Choi. The effects of manipulating task determinability on search behaviors and outcomes. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 445–454. ACM, 2018.

[14] R. Capra, J. Arguello, and Y. Zhang. The effects of search task determinability on search behavior. In *ECIR*, pages 108–121. Springer, 2017.

[15] A. Chaturvedi, P. E. Green, and J. D. Caroll. K-modes clustering. *Journal of Classification*, 18(1):35–55, 2001.

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

[17] M. J. Cole, C. Hendahewa, N. J. Belkin, and C. Shah. User activity patterns during information search. *ACM TOIS*, 33(1):1, 2015.

[18] J. S. Culpepper, F. Diaz, and M. D. Smucker. Research frontiers in information retrieval: Report from the third strategic workshop on information retrieval in lorne (SWIRL 2018). In *ACM SIGIR Forum*, volume 52, pages 34–90. ACM, 2018.

[19] A. Druin, E. Foss, H. Hutchinson, E. Golub, and L. Hatley. Children's roles using keyword search interfaces at home. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 413–422, 2010.

[20] D. Ellis. A behavioural approach to information retrieval system design. *Journal of documentation*, 45(3):171–212, 1989.

[21] M. J. Eugster, T. Ruotsalo, M. M. Spapé, I. Kosunen, O. Barral, N. Ravaja, G. Jacucci, and S. Kaski. Predicting term-relevance from brain signals. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 425–434, 2014.

[22] H. A. Feild, J. Allan, and R. Jones. Predicting searcher frustration. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 34–41. ACM, 2010.

[23] N. Fuhr. A probability ranking principle for interactive information retrieval. *Information Retrieval*, 11(3):251–265, 2008.

[24] J. He and E. Yilmaz. User behaviour and task characteristics: A field study of daily information behaviour. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, pages 67–76. ACM, 2017.

[25] C. Hendahewa and C. Shah. Implicit search feature based approach to assist users in exploratory search tasks. *Information Processing & Management*, 51(5):643–661, 2015.

[26] J. Huang, R. W. White, and S. Dumais. No clicks, no problem: using cursor movements to understand and improve search. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1225–1234. ACM, 2011.

[27] B. J. Jansen. Using temporal patterns of interactions to design effective automated searching assistance. *Communications of the ACM*, 49(4):72–74, 2006.

[28] B. J. Jansen and D. Booth. Classifying web queries by topic and user intent. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, pages 4285–4290. ACM, 2010.

[29] J. Jiang, A. Hassan Awadallah, X. Shi, and R. W. White. Understanding and predicting graded search satisfaction. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, pages 57–66. ACM, 2015.

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

[31] D. Kelly, J. Arguello, A. Edwards, and W.-c. Wu. Development and evaluation of search tasks for iir experiments using a cognitive complexity framework. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, pages 101–110. ACM, 2015.

[32] C. C. Kuhlthau. Inside the search process: Information seeking from the user's perspective. *Journal of the American society for information science*, 42(5):361–371, 1991.

[33] J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, pages 159–174, 1977.

[34] Y. Li and N. J. Belkin. A faceted approach to conceptualizing tasks in information seeking. *Information Processing & Management*, 44(6):1822–1837, 2008.

[35] J. Liu. Toward a unified model of human information behavior: an equilibrium perspective. *Journal of Documentation*, 73(4):666–688, 2017.

[36] J. Liu. Characterizing the stages of complex tasks. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1445–1445, 2019.

[37] J. Liu, M. J. Cole, C. Liu, R. Bierig, J. Gwizdka, N. J. Belkin, J. Zhang, and X. Zhang. Search behaviors in different task types. In *Proceedings of the 10th Annual Joint Conference on Digital Libraries*, pages 69–78. ACM, 2010.

[38] J. Liu, M. Mitsui, N. J. Belkin, and C. Shah. Task, information seeking intentions, and user behavior: Toward a multi-level understanding of web search. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, pages 123–132. ACM, 2019.

[39] J. Liu and C. Shah. Interactive IR user study design, evaluation, and reporting. *Synthesis Lectures on Information Concepts, Retrieval, and Services*.

[40] J. Liu and C. Shah. Proactive identification of query failure. *Proceedings of the Association for Information Science and Technology*, 56(1):176–185, 2019.

[41] Y.-H. Liu and N. Wacholder. Evaluating the impact of mesh (medical subject headings) terms on different types of searchers. *Information Processing & Management*, 53(4):851–870, 2017.

[42] I. Lopatovska. Toward a model of emotions and mood in the online information search process. *Journal of the Association for Information Science and Technology*, 65(9):1775–1793, 2014.

[43] M. Mitsui, J. Liu, N. J. Belkin, and C. Shah. Predicting information seeking intentions from search behaviors. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1121–1124. ACM, 2017.

[44] Y. Moshfeghi, P. Triantafillou, and F. E. Pollick. Understanding information need: An fmri study. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 335–344. ACM, 2016.

[45] A. Newell and H. A. Simon. *Human problem solving*, volume 104. Prentice-Hall Englewood Cliffs, NJ, 1972.

[46] E. Y. Rha, M. Mitsui, N. J. Belkin, and C. Shah. Exploring the relationships between search intentions and query reformulations. *Proceedings of the Association for Information Science and Technology*, 53(1):1–9, 2016.

[47] W. B. Rouse and S. H. Rouse. Human information seeking and design of information systems. *Information Processing & Management*, 20(1-2):129–138, 1984.

[48] S. Sarkar, M. Mitsui, J. Liu, and C. Shah. Implicit information need as explicit problems, help, and behavioral signals. *Information Processing & Management*, page 102069, 2019.

[49] R. Savolainen. Everyday life information seeking: Approaching information seeking in the context of "way of life". *Library & information science research*, 17(3):259–294, 1995.

[50] R. Savolainen. Cognitive barriers to information seeking: A conceptual analysis. *Journal of Information Science*, 41(5):613–623, 2015.

[51] C. L. Smith and S. Y. Rieh. Knowledge-context in search systems: Toward information-literate actions. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, pages 55–62. ACM, 2019.

[52] S. Stone. Humanities scholars: information needs and uses. *Journal of documentation*, 38(4):292–313, 1982.

[53] M. Swigon. Information barriers in libraries: types, typologies and polish empirical studies. *Library Management*, 32(6/7):475–484, 2011.

[54] K. Urgo, J. Arguello, and R. Capra. Anderson and krathwohl's two-dimensional taxonomy applied to task creation and learning assessment. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*, pages 117–124, 2019.

[55] P. Vakkari. A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *Journal of Documentation*, 57(1):44–60, 2001.

[56] Y. Wang and C. Shah. Investigating failures in information seeking episodes. *Aslib Journal of Information Management*, 69(4):441–459, 2017.

[57] R. W. White. Belief dynamics in web search. *Journal of the Association for Information Science and Technology*, 65(11):2165–2178, 2014.

[58] R. W. White. *Interactions with search systems*. Cambridge University Press, 2016.

[59] I. Xie. *Interactive information retrieval in digital environments*. IGI Global, 2008.

[60] I. Xie and C. Cool. Understanding help seeking within the context of searching digital libraries. *Journal of the American Society for Information Science and Technology*, 60(3):477–494, 2009.