A STATE-BASED APPROACH TO SUPPORTING USERS IN COMPLEX SEARCH TASKS

by

JIQUN LIU

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Communication, Information, and Media

Written under the direction of

Chirag Shah

and approved by

New Brunswick, New Jersey

May, 2020

ABSTRACT OF THE DISSERTATION

A State-Based Approach to Supporting Users in Complex Search Tasks

by Jiqun Liu

Dissertation Director: Chirag Shah

Previous work on task-based interactive information retrieval (IR) has mainly focused on *what* users found along the search process and the *predefined, static* aspects of complex search tasks (e.g., task goal, task product, cognitive task complexity), rather than *how* complex search tasks of different types can be better understood, examined, and disambiguated within the associated multi-round search processes. Also, it is believed that the knowledge about users' cognitive variations in task-based search process can help tailor search paths and experiences to support task completion and search satisfaction. To adaptively support users engaging in complex search tasks, it is critical to connect theoretical, descriptive frameworks of search process with computational models of interactive IR and develop personalized recommendations for users according to their *task states.* Based on the data collected from two laboratory user studies, in this dissertation we sought to understand the states and state transition patterns in complex search tasks of different types and predict the identified task states using Machine Learning (ML) classifiers built upon observable search behavioral features. Moreover, through running Q-learning-based simulation of adaptive search recommendations, we also explored how the state-based framework could be applied in building computational models and supporting users with timely recommendations.

Based on the results from the dissertation study, we identified four intention-based task states and six problem-help-based task states, which depict the active, planned dimension and situational, unanticipated dimension of search tasks respectively. We also found that 1) task state transition patterns as features extracted from interaction process could be useful for disambiguating different types of search tasks; 2) the implicit task states can be inferred and predicted using behavioral-feature-based ML classifiers. With respect to application, we built a search recommendation model based on Q-learning algorithm and the knowledge we learned about task states. Then we apply the model in simulating search sessions consisting of potentially useful query segments with high rewards from different users. Our results demonstrated that the simulated search episodes can improve search efficiency to varying extents in different task scenarios. However, in many task contexts, this improvement often comes with the price of hurting the diversity and fairness in information coverage.

This dissertation presents a comprehensive study on state-based approach to understanding and supporting complex search tasks: from task state and state transition pattern identification, task state prediction, all the way to the application of computational state-based model in simulating dynamic search recommendations. Our process-oriented, state-based framework can be further extended with studies in a variety of contexts (e.g., multi-session search, collaborative search, conversational search) and deeper knowledge about users' cognitive limits and search decision-making.

Acknowledgements

First and foremost, I would like to express my deep and sincere gratitude to my mentor and advisor, Professor Chirag Shah, for his guidance, encouragement, patience and support throughout my entire PhD journey. I am grateful to him for opening me to the exciting world of information retrieval research and pushing me to pursue research excellence through his rigorous training. More importantly, he never fails to demonstrate his faith in my work and research potential when I got rejections from journals, conferences, and funding agencies. Also, as an introvert, I am grateful to him for teaching me how to interact with people and enjoy social gatherings and activities in academia. I am not the kind of person that everyone likes (perhaps never will be). However, he shows me the ways of being a nice person to people around me and also enjoying life outside the world of research and science. I am fortunate to have him as my research advisor and role model in life.

I would like to thank Professor Nick Belkin, for his invaluable guidance, encouragement and support throughout all stages of my PhD process. I am grateful to him for offering me the opportunities to work in his projects and participate in the cutting-edge research on interactive information retrieval. I am also grateful to him for being a great mentor and training me for my future research life.

I would like to thank my committee members, Professor Kaitlin Costello and Dr. Daniel Russell for their advice on my dissertation work. I thank Professor Costello for offering insightful comments on my research design and theoretical framework, which pushes me to explore deeper on my research problem both theoretically and empirically. Dr. Daniel Russell also provided immense support and critical comments on my dissertation work. Despite his busy schedule, he always asked detailed, critical questions via our emails and zoom meetings and provided word to word comments on my dissertation writing. All of these comments, questions, and encouraging advice helped me improve and solidify my dissertation work presented here. Thank you.

My dissertation research has also benefited from the insightful comments and questions from researchers at doctoral consortia and academic conferences. They include: Professor Diane Kelly, Professor Gareth Jones, Professor Noriko Kando, Professor Jacek Gwizdka, Professor Katriina Byström and Professor Jaime Arguello.

I would like to thank the professors with whom I've interacted extensively in the School of Communication and Information and Department of Computer Science. Without their help and support, this PhD certainly would not have been possible. I apologize for missing anyone: Professor Vivek Singh, Professor Nina Wacholder, Professor Marie Radford, Professor Charles Senteio, Professor Lilia Pavlovsky, Professor Michael Doyle, Professor Jennifer Theiss, Professor Xiaojun Yuan, and Professor Yongfeng Zhang. Also, I would like to thank Professor Ying-Hsang Liu from University of Southern Denmark and Professor Anna Maria Tammaro from University of Parma, for their friendship, encouragement and faith in my research potential.

I would like to thank members of the InfoSeeking laboratory and SC&I PhD cohorts for their assistance and friendship throughout the years, and again I apologize for missing anyone: Yiwei Wang, Souvick Ghosh, Soumik Mandal, Dr. Matthew Mitsui, Shawon Sarkar, Manasa Rath, Ruoyuan Gao, Diana Floegel, Jonathan Pulliza, Dr. Dongho Choi, Isha Ghosh, Jiho An, Shannon Taber, SeoYoon Sung, Wei Shi, Eun Youp Rha, and Deborah Yoon.

I would like to thank my professors and friends at Nankai University and Peking University, who encouraged me to apply for PhD program and pursue my research career: Professor Hui Yan, Professor Yuelin Li, Professor Chang Liu, Professor Pengyi Zhang, Professor Qingshan Zhou, Yufei Shen, Yang Sun, Yi Bu, and Jiahui Wu.

I would also like to thank National Science Foundation (NSF) Awards #1423239 and #1717488, which have provided financial support for the research studies in the information seeking intention and problem-help projects.

And lastly, I would like to thank myself for never giving up when facing all sorts of difficulties in research and life.

Dedication

This dissertation is dedicated to my wife, Jixin Lin, and my parents Shujun Fang and Shejiao Liu. Thank you for your unconditional love and support.

Table of Contents

List of Tables

List of Figures

Chapter 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 [11, 12]. Search systems and technologies have experienced phenomenal success in recent years, especially in addressing fact-finding and navigational search tasks [141]. However, current search systems, task and interaction models, and the underlying algorithms still face plenty of challenges when applied in supporting *complex tasks* which involve multi-round, multidimensional search interactions (e.g., planning a research project, comparing and evaluating retirement packages) [6, 35]. 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. Most of the previous research attempts on this problem can be roughly grouped into two categories: (1) One-dimensional approach: researchers define tasks according to their positions/values in one single dimension or spectrum, such as cognitive complexity (e.g., [69, 148]) and task prior determinability (e.g., [28, 27]). (2) Holistic/multi-dimensional approach: researchers propose and implement a holistic, multifaceted framework which defines a task as a combination of multiple facet values [78, 84]. Both of these approaches have theoretically supported a large body of IIR research concerning identifying and responding to users' tasks as static, overarching goals or problems that drives search iterations and explorations of potential solution spaces. However, very little research has attempted to explore how tasks and problematic situations are unfolded and evolve during the process of search interactions, and how we can personalize and optimize search system supports according to the dynamic states of tasks and user characteristics (e.g., information seeking intentions, local search tactics, problems encountered during search).

In addition to the multidimensionality discussed above, task is also a multilevel concept. Specifically, task as a concept of 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 [25]. From a process-oriented perspective, the sequence and transitions of states in a search session at multiple levels often reveal essential properties of a search task as well as the task doer (e.g., search skills, knowledge base). In the light of Newell and Simon's human problem-solving framework [103], the transitions of states and actions can be considered as representations of users' iterative explorations in the evolving solution space behind the problem associated with the search task at hand.

Integrating human problem-solving perspective with the static definitions of task complexity, we define *complex search tasks* 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:

- Predefined search goal(s) that emerge from the related motivating task.
- Users' search skills, topical and procedural knowledge regarding the task at hand, and the internal structure of information processing systems (IPS).
- Unanticipated results and search problems.
- Affordances and supports from the available interactive search system(s) at different moments of search.

In contrast to the static view of task complexity (e.g. [69, 27]), the process-oriented definition of complex search task is built upon an underlying 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. When talking about the complexity of a search task, we are essentially talking about the complexity for a particular user or a group of users to do a search task with the support from a given search system (e.g. commercial search engine).

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 [88], cognitive biases [140], behavioral states in search session segments [54], encountered problems [137], level of search satisfaction [64] and other aspects of search interactions. In order to explore this in greater depth, this dissertation study aims to (1) reveal and characterize the dynamic nature of complex search tasks from a process-oriented perspective, and then (2) leverage this knowledge in supporting users during task-based search sessions. Note that although the main focus of this dissertation is on characterizing and supporting complex search tasks in users' interactions with existing search systems (in this case, commercial search engine), the ultimate goal of this line of research (for which the findings from this dissertation study could be an appropriate point of departure) is to (1) understand searchers' motivating work tasks and problematic situations and to (2) identify, design, and evaluate possible system actions and techniques for dynamically supporting different task states as well as users' explicit and implicit needs.

1.1 The context and approach of the dissertation

The studies of tasks and associated search interactions are mainly situated in the broad research area of information seeking and searching. Information seeking behavior is the purposive seeking for information as a consequence of a need to accomplish some goals or resolve a problematic situation [11, 145]. Within the scope of information seeking, information searching behavior is the totality of behavior employed by the searcher when interacting with information systems of all kinds [145, 146]. Despite the diversity of the specific theories of interactive information seeking and searching behavior, most of these theories have been developed to address two common research problems: (1) How people behave when seeking and searching for information? (2) What are the motivations and contexts (e.g., motivating task) behind information seeking and search activities? To answer these questions, two central aspects of information seeking and searching need to be analyzed both conceptually and empirically: (1) the components, patterns, and outcomes of information seeking and searching behaviors and (2) the components of the contexts in which these behaviors occur, including the components from both personal contexts (e.g., conceptual state of knowledge, cognitive gap in sense making, emotional state) and environmental contexts (e.g., information scent level and pattern, task facets, problematic situation, available supports from systems and people). Thus, most, if not all, of the existing information seeking and IIR models have been developed to describe and explain at least some of the facets related to one or both of these two central aspects.

As a natural extension of the varying approaches to the two central questions above, many information seeking and IIR researchers have proposed multiple ways to leverage the knowledge of users' search interactions (e.g., implicit relevance feedback) and the associated contexts (e.g. task) in satisfying information needs and facilitating task completion. For instance, Liu et al. [81] developed a method for modeling search behaviors to predict document usefulness and then using the knowledge of users' behavioral patterns to personalize search results in tasks of different types. Yuan and Belkin [156] demonstrated the effectiveness of an integrated IIR system which adapts to support different information seeking strategies by comparing it with a standard baseline IR system in a lab experiment. With respect to task-oriented supports, Ahn et al. [3] developed a Web search system that can extract relevance feedback based profile (i.e. task model) from search interactions and utilize the profile for search result personalization.

When defining tasks as overarching goals that drive information seeking and searching activities, it is difficult to reconcile the conflict between static task properties (e.g., task product, task goal) and the dynamic nature of search interactions. This difficulty leads to at least two fundamental challenges: (1) In a task-based search session, the varying states (especially users' task-related cognitive states) behind the transitions of search tactics are not clear; (2) Instead of merely offering task-level static recommendations, a search system needs to implement an explainable state-based approach and to dynamically support users at different moments of task completion and problem solving processes. These two fundamental challenges contextualize and motivate the dissertation work reported here.

1.2 Problem statement

From a process-oriented perspective, a complex search task can be represented as a sequence of different states that are connected through users' search actions. Note that a state is fundamentally different from a *subtask*. The implicit assumption behind the notion of subtask is that the underlying overarching task can be planned *before search* and be decomposed into separate smaller predefined steps or atomic information needs [96]. On the contrary, a user's current task state is affected by his or her previous task state, previous search experience and judgments, elements of current local situation (e.g., search intentions, problems encountered), as well as the static aspects of the overarching search task and the associated motivating task (e.g., task product, task goal, the user's knowledge structure). Therefore, under the influence of the situational aspects of search (e.g., local search experience and performance, type of situation, adaptive system supports), users' task states often change and evolve during search interactions and cannot be predefined as static subtasks of overall task-based workflows or simply be represented by the planned aspects of search interactions [151].

The state-based, process-oriented perspective offers us an alternative approach to answering the central research questions and to dynamically supporting users. Specifically, to answer the two central questions (i.e. describe search behavioral patterns and explore the underlying motivations and contextual factors) in information seeking and IIR communities, we need to address the following two problems/questions:

- What are the states of complex search tasks and state transition patterns in tasks of different types?

- What is the relationship between observable search actions and implicit task states?

To devise scalable recommendations, we need a computationally-congenial, state-based task representation that can support the simulation and evaluation of adaptive search system actions and the associated rewards. Thus, built upon our study on the first two research problems, we aim to address an additional application-oriented research problem:

- How can a search system provide generalizable, scalable state-based recommendations to support users engaging in complex search tasks?

The following sections and the *Literature Review* chapter will provide more detailed discussions regarding the research background that contextualizes the above research problems as well as the potential solutions to these problems.

1.3 Dynamic modeling: States of complex search tasks

From a process-oriented perspective, when users engage in information seeking and searching episodes, they are essentially exploring, attempting, and evaluating one or more alternative paths in order to reach the goal states that are determined by both a set of desired (specific or amorphous) outcomes and ongoing interactions with information and systems. If we conceptualize task-based search interaction as a human problem-solving process (cf.[103]), then the states of a given search task for a user can be considered as a set of possible positions distributed in an *accessible* solution space. In this sense, different search paths can be represented by different sequences of task states. For complex search tasks, not all search paths in accessible solution space can finally lead to the goal/desired state. The future (both next-step and final) performances of alternative solutions are usually shaped by past search iterations and thus need to be examined during search sessions.

In addition to the variations in states, search actions, and ongoing search paths, another dynamic aspect of search tasks is that the desired goal states often vary due to the changes in users' task perception (e.g., perceived task difficulty [4, 82]), knowledge of the task and associated topic(s) [88, 160], expectations and thresholds of satisficing [1, 89, 87], and other contextual factors (e.g., offline information sources and supports, interruptions and costs of task resumptions). Specifically, for example, after a series of failed attempts (i.e. queries and search result examinations) in addressing some parts of a given task, a user may consider these parts as unsolvable in the context of search and hence may remove them from the final goal state and search satisfaction criteria. This dynamic aspect of desired state, along with the variations in the sequence of states and actions, demonstrate the value and necessity of developing a dynamic framework of complex search task and reinforce the idea of implementing a state-based approach in supporting users' interactions with search systems.

Depending on the specific dimensions and perspectives, there are multiple ways of representing and measuring the states of tasks, which may contribute to the understanding of different aspects of task-oriented problem-solving processes. On the *active, intentional* dimension, task state at a given moment of search can be defined as the things that a user is trying to accomplish within a search iteration/query segment. A query segment refers to a search session segment which starts from one issued query and ends at the next query [100]. Each individual thing or local goal that motivates a query segment is defined as an information seeking intention [88]. Since users usually try to accomplish multiple things within each individual query segment, their task states can be represented by the combinations of individual information seeking intentions (i.e. intention states). From this perspective, the dynamic aspect of complex search tasks can be at least partially captured through studying the transitions between various intention states and the associated search tactics in query segments.

On the *situational, unplanned* dimension, task state can be represented by the problem(s) that a user encounters at different moments during a search session and the type(s) of system help and support that the user prefers to have for addressing the problem(s). In this sense, a user's task-based search interaction can be conceptualized as a continuous problem-solving process where the user explores potential solution space by addressing the emerging problems along the way. The problem(s) encountered at each moment is determined by the past problems as well as the user's search tactics. The problem-solving process stops when (1) the user reaches the desired goal state, or (2) the user is frustrated and decides to abandon the search session.

Based on the definitions of task state on intentional and situational dimensions, this dissertation study adopts a process-oriented perspective and seeks to represent complex search tasks using the sequences of task states as well as the associated search actions which push the transitions of states.

1.4 Dynamic support: State based system recommendations

Recall that a large body of existing information seeking and IIR studies sought to address the two central questions/problems in the research community: (1) how people seek and search for information (2) what are the underlying motivations and contexts (e.g., motivating task). One of the ultimate goals of advancing knowledge regarding these two central questions is to design and evaluate interactive search systems that can leverage the knowledge of users' motivations and contexts in supporting search interactions. Many empirical studies demonstrate that personalized search systems that exploit the user's motivating task features can significantly improve the effectiveness of information retrieval (e.g. [3, 70, 143, 81, 153]).

To achieve effective personalization, a search system needs to know not only *what* are the things that a user is trying to do (the predefined goals and static aspects of tasks [25]), but also *how* they are doing them (the changing states and the adaptation in search tactics [34]). Based on the state-based approach discussed above, this dissertation plans to go beyond the predefined, static task aspects and take advantage of the knowledge of users' dynamic states, actions, and feedbacks in personalizing search recommendations. Furthermore, we seek to explore the methods of automatically inferring users' implicit states from multiple explicit signals in search interactions and thereby providing state-based recommendations in wider contexts without worrying about the availability of state annotation data. More details regarding the state-based search path evaluation, recommendation and state inference methods are explained in the *Methodology* chapter.

1.5 The study overview

To address the proposed research problems, the dissertation first develops state-based models of complex search tasks. The information about users' task states is extracted from users' post-search annotations in user studies. On the *active* dimension, the central component is the combination of individual information seeking intentions. On the *negative* dimension, the core element includes the problem(s) that a user encounters and the help(s) that the user wishes to have for addressing the problem(s). Analyzing the state annotation data can reveal the structure of task process on different dimensions. We argue that in the context of information seeking and searching, a task is not merely a set of static, predefined goals waiting to be reached. Instead, a task also contains user-related, transitory elements which can be represented by the distributions and transitions of task states and users' actions.

Users' annotations of information seeking intentions, encountered problems, preferred help and the associated search behavioral data are from two controlled lab studies, information seeking intention study (ISI) and problem-help study (PH). The ISI study collected intention annotations and search behavior data from 693 query segments generated by 40 participants in 80 task-based search sessions. In the PH study, a total of 273 query segments were logged, along with users' annotations on perceived search problems and preferred help.

This dissertation study uses users' annotations as rich representations of their task states in multiple-round search iterations. In addition to exploring the associations between task states and search behaviors, we also seek to develop practical methods that can generate empirically-grounded, state-based recommendations of different types for improving users' search effectiveness at multiple levels.

More details regarding the user study designs, datasets, and analytical algorithms are provided in the *Methodology* chapter.

1.6 The knowledge gap addressed by the dissertation

When users search for information with the goal of completing complex search tasks, they usually experience varying task states and change their search tactics due to the variations in the information space created by search results and interface layouts, knowledge and perception of the tasks, expectations regarding system responses, as well as other situational factors. For most complex or difficult searches, the transitions of states and adjustments of search tactics cannot be planned in advance mainly because of the difficulty in predicting informational gains and the unexpected obstacles and opportunities in search process. Through building, implementing, and evaluating state-based models of complex search tasks, this dissertation seeks to advance knowledge in the following three aspects:

- Representing and explaining the dynamic aspect of complex search tasks using the distributions and transition patterns of task states.
- Revealing the connection between task states and users' search tactics and thereby paving the way for building state-aware adaptive search supports.
- Leveraging the knowledge of users' task states in improving and personalizing search recommendations at different moments of search.

1.7 Summary

This chapter presented the problem statement of this dissertation work, and explained the motivations behind why advancing knowledge on the specific research topic, states of complex search tasks, is of importance. The problem we are addressing is the lack of approaches to provide dynamic system assistance to users experiencing varying states of tasks in complex search sessions. The conceptual issue behind this problem is that the static definition of tasks (i.e. a set of planned, predefined desired goals) cannot fully explain the process-oriented, dynamic aspects of tasks [128]. Users' information seeking and search strategies are the products of both planned aspects and dynamic situations [151]. This chapter also provides an overview of the proposed approach, which will be discussed in detail in the methodology section.

The next chapter presents a literature review concentrating on multiple facets and levels of tasks in the context of information seeking and searching. Then, a state-based theoretical framework is constructed based on recent research progress on information seeking intentions, users' search problems and supports needed, as well as the existing theories developed based on the empirical evidences from a large body of task-based information seeking and IR studies. To inform the implementation of the state-based approach in simulating and evaluating search recommendations, our review also includes some of the recent research on personalized search recommendations. Lastly, within the state-based theoretical framework, specific research questions for this dissertation are proposed based on the problem statement in this *Introduction* chapter.

Chapter 2

Literature Review

This chapter provides a review of research in which the dissertation is situated. The main relevant research areas include: Information Seeking, Interactive Information Retrieval, Search Recommendation and Personalization, and Reinforcement Learning (especially Markov decision process). Reviewing and discussing previous works in these areas paves the way to understanding (1) why this dissertation study is of importance, (2) what are the available theoretical and empirical supports from existing research, and (3) in which research area(s) we are advancing the knowledge.

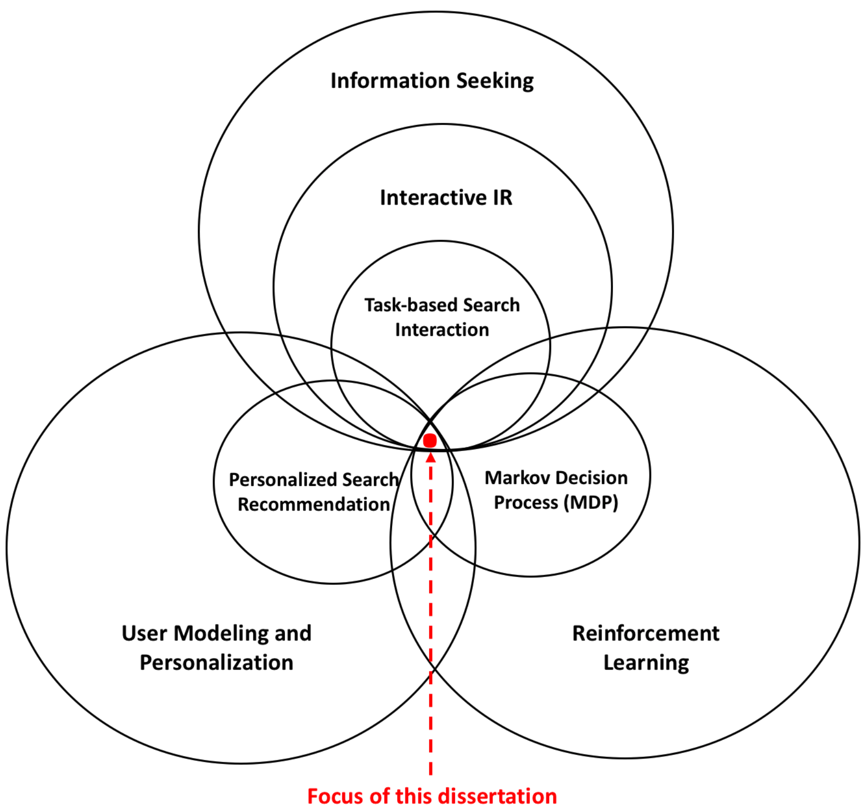


Figure 2.1: Position of the proposed dissertation study

To answer these three questions, it is important to first clarify the position of our work within broader, interrelated research communities. Figure 2.1 illustrates where the proposed dissertation study lies at the intersections of the relevant domains. More detailed discussions on previous relevant studies are presented in the following sections.

The literature review plan starts with looking at the task construct in the context of information seeking and searching and analyzing its different levels, aspects and dimensions. Review of task properties includes both static facets and dynamic aspects (i.e. process or stage models of tasks). In particular, this section also summarizes the recent research progresses on understanding, modeling, and supporting complex search tasks and highlights the new opportunities of implementing state-based approach. The next section explores the local contexts and states within task-based search sessions. Specifically, we discuss the past research on both active side (i.e. information seeking intentions) and passive side (i.e. encountered problems and search barriers) in search iterations. The review then turns to the discussion of search behavior analyses with a focus on the widely adopted behavioral constructs (e.g., tactics, strategies, paths) and the associated behavioral measures. Finally, some frameworks and methodologies of supporting and personalizing search interactions are reviewed. We discuss Markov Decision Process (MDP) approach in detail and explain its suitability to the problem of leveraging the knowledge of task states and dynamically supporting users engaging in complex search tasks.

2.1 Tasks in information seeking and searching

In much of previous information seeking and IIR research which involves the concept of task, a task was conceptualized either as an abstract representation of a set of desired goals (static perspective) or as a concrete sequence of states and actions (dynamic perspective) [25]. The way in which we define and operationalize tasks can significantly affect both our understanding of users' search strategies and our approaches of designing and evaluating interactive search systems. A large body of empirical studies has explored the static, objective aspects of tasks (e.g., task product, task goal) and sought to understand, measure, and predict these task facets from content-based features and search behavioral measures. However, little research has examined the process of search tasks and the implicit variations in some of the task properties (e.g. intention states, task-related perceptions and expectations) during search interactions.

To develop a deep, comprehensive understanding of the task construct, we first look at different levels of tasks and explain the possible connections among these levels. Furthermore, to shed light on the broader context of tasks, this section also includes a brief discussion of the potential beyond-task impacts of human information interaction. Then, we review several widely-studied facets of tasks in past information seeking and IIR studies. This section then turns to the process models of tasks, aiming to reveal the dynamic aspects of tasks. Particularly, the recent studies on understanding and supporting complex search tasks are discussed in depth.

2.1.1 Task levels

According to Byström and Hansen [24, 25], task contexts in information practices can be represented by a nested model consisting of three levels (from outer level to inner level): work task, information seeking task, and search task (see Figure 2.2). Specifically, work tasks are separable parts of a person's duties in his or her workplace [25]. Note that not every subtask within a work task can be transformed into an information seeking task. In many cases, some parts of a work task need system and human supports that are beyond the capacity of search systems (e.g., writing a dissertation proposal). In addition to the tasks generated in workplaces, everyday life tasks that emerge from non-work scenarios can also lead to active information seeking and searching practices (e.g., search for and book a hotel for travel) [2].

Information seeking tasks are a central component of information-intensive work tasks and may be deconstructed into general stages, including task construction, task performance, and task completion [24]. To identify the implicit information seeking task(s) within a work task, people need to analyze the information that is needed as well as the availability of various information resources and supports. This analysis is influenced by both work task properties and the task performer's knowledge and experience of using information resources.

Information search tasks focus on the satisfaction of a separable fraction of an information need through a single consultation of a source or sources (especially search systems) [25]. Many facets of a search task are significantly affected by the corresponding properties of the overarching work task [77]. Due to the integration of search supports and general artificial intelligence, many recently developed intelligent systems (e.g., Google assistant, Alexa) seek to go beyond simple search tasks and to directly support actions of different types in information-intensive work tasks (e.g., estimate task duration and arrange schedules [142], provide conversational guided task supports [136]).

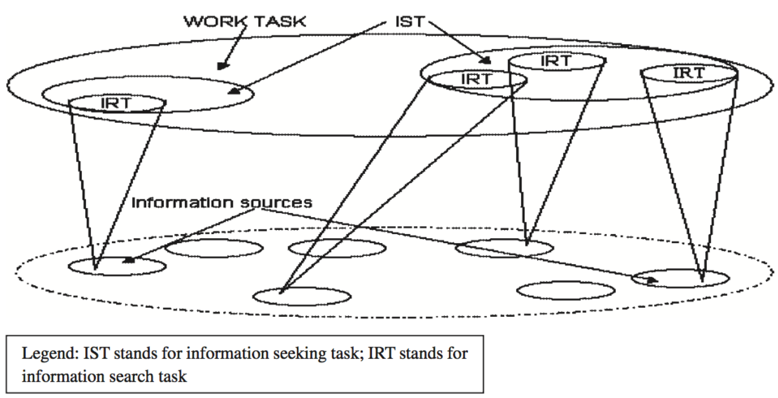


Figure 2.2: Byström & Hansen's model of task levels [24].

In addition to Byström and Hansen's nested model of task [24], Xie [151] also explored the multilevel nature of user goals and tasks and developed a four-level hierarchical framework of goals, including long-term goals (e.g., users' personal interests), leading search goals or work tasks, current search goals (current information seeking and search tasks), and interactive intentions (things that a user wants to accomplish in local steps or stages of search). This four-level typology covers a wide range of user goals and tasks (from long-term task-independent goals to local goals behind specific search tactics) and was verified by a series of user studies [151, 80].

To fully understand the role of tasks, it is also important to explore the value and impacts of information seeking and searching that go beyond specific task contexts. Although many information seeking episodes are driven and shaped by tasks of different levels, it does not mean that the influence of information seeking and searching is restricted within the immediate task contexts. Instead, many information seeking and search tasks actually serve as the opportunities for users to enhance the metacognitive, information-literate skills that are often required for long-term learning and critical thinking [124] and to adjust their respective cognitive spaces and images of the external world [60]. Therefore, when conceptualizing and examining tasks in human information interaction, researchers need to consider both task-centric factors and learning-centric elements. The investigation of different levels of tasks (including the long-term, learning-oriented aspect that goes beyond immediate task scenarios) can generate distinct focuses and metrics for search system evaluations.

2.1.2 Task facets

To gain a comprehensive understanding of the impacts of tasks on information seeking and search behaviors at multiple levels, information seeking and IIR researchers have explored a variety of task dimensions or facets and sought to classify tasks on the basis of one or multiple dimensions. Focusing on different dimensions or task taxonomies, previous research has examined the impacts of task facets on search interactions from different perspectives. For instance, Liu et al. [84] and Jiang, He, and Allan [65] examined the associations between user behaviors and objective task features (i.e., task product, task goal, task type) and discussed to what extent these behavioral features can help disambiguate search tasks of different types. Capra et al. [27] found that manipulating task a priori determinability via modifying task items and dimensions can significantly affect users' perceived task difficulty and choices of search strategies.

With respect to task-user combined features, Wildemuth [144] argues that in task-based information search, users' search tactics are influenced by their domain knowledge related to task topics. Liu, Gwizdka, Liu, and Belkin [85] demonstrates that both whole-session level and within-session search behaviors are affected by task difficulty, and that the dynamic relationships between search behavior and task perception are subject to the influence of task type (i.e., single fact-finding, multiple fact-finding, and multiple-piece information gathering). Similarly, Aula, Khan, and Guan [5] also investigated search behavioral variations under tasks of different levels of difficulty. By conducting a lab study and a large-scale online study, they found that when performing difficult search tasks, users tend to issue more diverse queries (have a more unsystematic query refinement process), use advanced operators more frequently, and spend longer time on search engine result pages (SERPs) during their search processes.

Given that many IIR studies only examine one or a few task dimensions, Li and Belkin [78] developed a faceted approach to conceptualizing tasks in IR based on related literature on task classification as well as their empirical studies on task-based information searching [77, 79]. The faceted framework provides a holistic approach to exploring the nature of tasks and conceptually supports a series of empirical studies on task-based search interactions.

2.1.3 Process models of tasks

Task process is one of the facets of the task entity [78]. Differing from the static task properties (e.g., predefined task goal, task product), however, task process speaks to an alternative approach to understanding the nature of tasks. When conceptualizing tasks from the process-oriented perspective, we are essentially looking at the process of *doing tasks*. The core argument behind this perspective is that in the context of information seeking and searching, we cannot define or study a task without examining *how* the task was actually completed (or failed). Therefore, to fully understand a task, we need to explore both the objective task features and users' responses to the evolving task environments at multiple levels (e.g., behavioral, cognitive, emotional).

In information seeking and IR communities, a series of classical models have been developed and applied to describe the general process of performing information seeking and search tasks. Many of these models mainly focus on the behavioral aspect of task process and look at the transitions of information seeking and search actions. For instance, to describe the general process of information seeking, Ellis [41] studied the information seeking pattern of academic social scientists and broke it down into six characteristics: starting, chaining, browsing, differentiating, monitoring, and extracting. Wilson [146] suggests that in some circumstances, Ellis's characteristics can be organized as a sequence of information seeking stages in a process model. Ellis's model clearly identifies the features of information seeking patterns and has been modified and tested based on empirical studies (e.g., [42, 43]). However, this model only describes the behavioral level of task-based information seeking. It does not touch the interaction between the information seeker and the multi-dimensional context in which task states and information seeking activities evolve.

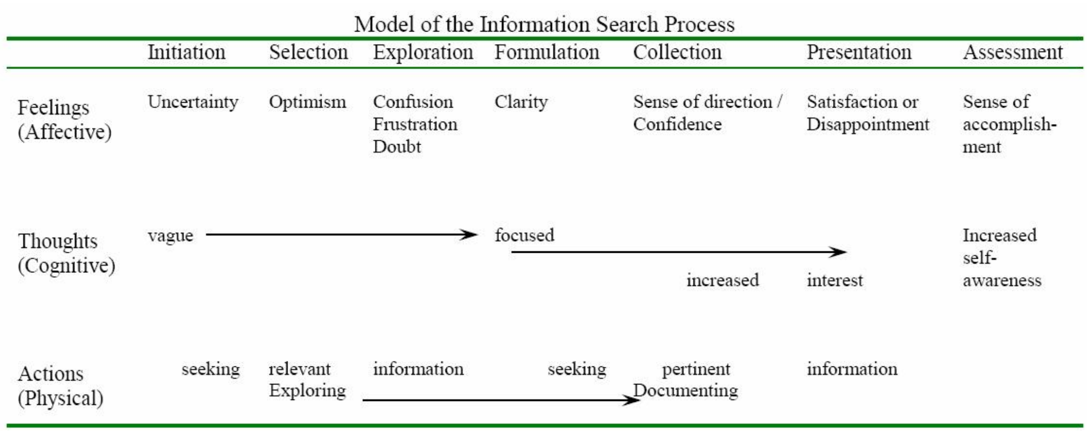


Figure 2.3: Kuhlthau [72]'s Information Search Process (ISP) model.

Given the multidimensionality of information seeking activities, Kuhlthau's work complements that of Ellis by attaching to the stages of the "Information Search Process" (ISP) the associated affective states (e.g., uncertainty, sense of direction), cognitive states (e.g., vague, focused) and actions [72, 73] (see Figure 2.3). Similar to Ellis's framework, Kuhlthau's ISP model has also been tested in many empirical studies conducted in library and educational contexts (e.g. [8, 72, 74]). In addition, Kuhlthau [73] also proposes the principle of uncertainty that states that information commonly increases uncertainty in the early stages of the information seeking and search process. The increased uncertainty indicates a space for interactive systems to provide in-situ interventions and task-centric supports. The concept of uncertainty in ISP connects the affective level and the action level of information seeking and is also associated with certain cognitive states and problems in human-information interaction, such as anomalous state of knowledge in interactive information retrieval [9] and cognitive gap in the Sense-Making process [36]. Kuhlthau's ISP model is a useful tool for describing and qualitatively explaining the stages of information seeking at multiple levels.

However, it offers limited insights for understanding the variations in task states and behavioral changes within information search sessions. This is because the ISP model was developed in the context of long-term learning and information seeking tasks, and described the stages over month(s)-long periods rather than individual search sessions.

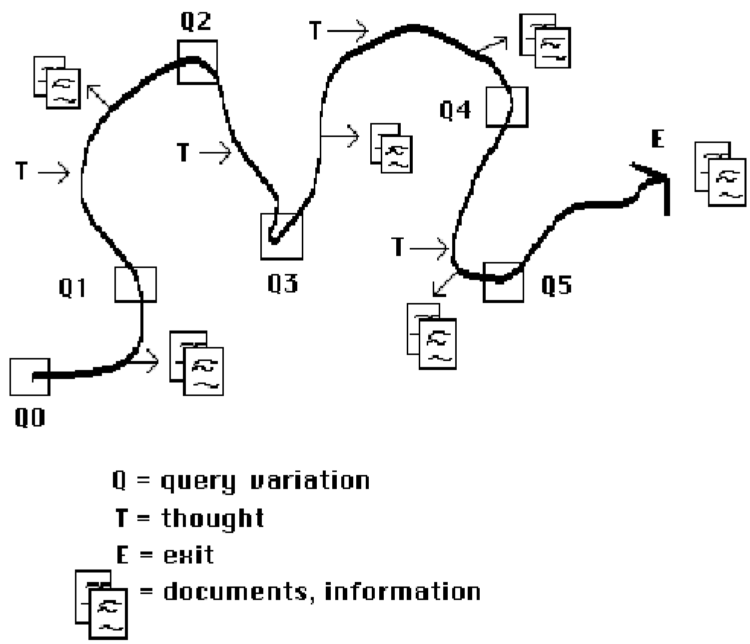


Figure 2.4: Bates [7]'s model of berrypicking, evolving search.

In the IR community, a number of models and techniques have been developed to describe and explain different aspects of tasks and search activities. Oddy [104] developed the THOMAS program for supporting users' dialogues with IR systems. The THOMAS model takes into account the shift in users' needs and emphases during search interactions and adjusts its information displays according to users' needs and reactions or judgments to the retrieved information. In contrast to the relevance feedback model which assumes that users' information needs are static (the only improving part is query formulation), the THOMAS model offers more room for users to express and shift their focuses during search iterations and the associated learning processes. Similarly, Bates [7] proposes the berrypicking model to describe the process of information searching. Specifically, Bates [7] argues that the classical single-query, best-match model cannot capture the interactive, evolving nature of information search tasks (see Figure 2.4). In the berrypicking model, the nature of query is an evolving one, instead of single and static. Also, the search process follows a berrypicking, evolving process, rather than a linear sequence of steps leading to a single best retrieved set of relevant documents. In contrast to the traditional single-query model of ad hoc information retrieval, the berrypicking model illustrates the interactive process of information searching and has been empirically supported by many task-based studies in the information seeking community [41, 113, 127, 72]. Spink [126] developed a multi-level model of search and identifies user judgments, search tactics, interactive feedback loops, and cycles as constituting the search process of an IR system user in tasks of different types. Based on Kuhlthau's ISP model as well as a series of empirical studies, Vakkari [135] proposes a general framework of task-based information searching which consists of three task 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. Belkin [10] proposed a conceptual model that represents session-level information seeking episodes as a sequence of users' iterative interactions with an interactive search system and the retrieved information objects (e.g., search result snippets, documents) (illustrated in Figure 2.5). The focus of Belkin [10]'s search session model is a user-centered search interaction that varies over time under the influence of task(s), goal(s), and an evolving problematic situation. Although Belkin [10] did not specify the states and state transition patterns in a search session, he clearly emphasized the basic ideas of modeling temporal changes in search interactions and developing dynamic supports for users according to search task properties.

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 designing practical, dynamic supports for complex tasks at different moments. To address this issue, some researchers seek to develop computationally-congenial models for representing task states or stages, simulating task-based search interactions, and

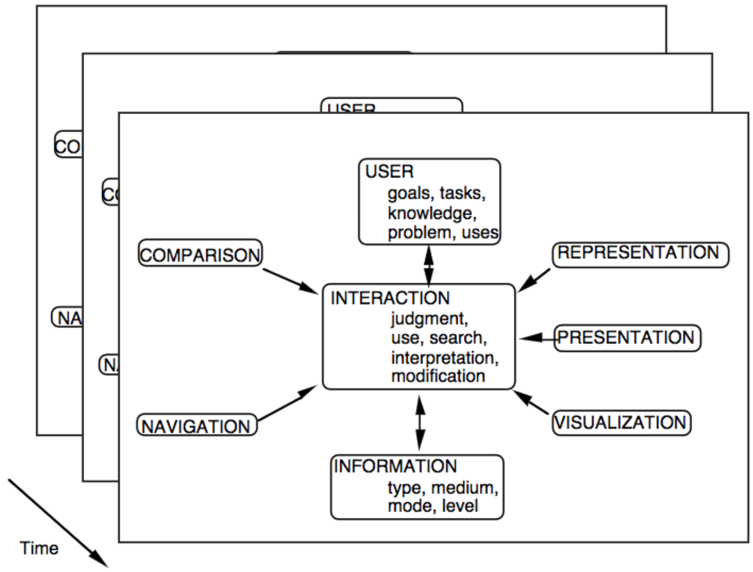


Figure 2.5: Belkin [10]'s model of user-centered search interaction

evaluating recommendations of different types. For instance, Cole et al. [34] investigated user activity patterns in tasks of different types and demonstrated that task types and levels of task difficulty can be represented and disambiguated by the sequence and distributions of user activity states (derived from page visiting behaviors) and cognitive processing states (approximated using eye movement patterns). Similarly, Dung and Fuhr [38] adopted both discrete and continuous search behavioral signals and develops a Hidden Markov Model (HMM) for recognizing the search phases and analyzing the transitions among them. To develop an effective formal model of search interactions, Fuhr [46] proposed a framework for extending probabilistic IR to the IIR context and representing users' situation transitions and choices at different moments of information searching episodes. He found that within the proposed cost model of interaction, the expected benefit of a single choice can be maximized, which forms the basis for the derivation of the optimum ordering of choices (i.e. the probability ranking principle for interactive IR, or IIR-PRP). The IIR-PRP model serves as an important step towards building a computational framework for supporting the functional design of interactive search systems. However, this model abstracts out a variety of user characteristics and lacks effective representations of the task states and associated cognitive variations.

2.1.4 Complex search tasks

As it is discussed in the previous sections, an important area of IIR research involves understanding and measuring the impacts of task facets on search behaviors, experiences, and performances. *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 [26] 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 require different types of information through a variety of information sources at different points of information seeking epsides. Kelly et al. [69] explored the cognitive complexity of task and adopted Bloom [17]'s taxonomy of learning domains in characterizing different levels of search task complexity. Their results indicate that complex search tasks (e,g., analyze, evaluate, create) required significantly more search activities from users (e.g., more issued queries, clicks, and dwell time on visited pages and documents). Ghosh, Rath, and Shah [49] employed Kelly's framework of task complexity in their user study on searching in learning-oriented tasks. They found that tasks of different cognitive complexities varied significantly in information search patterns (the transitions of search tactics) and learning performances. 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 can significantly affect the overall task determinability [27, 28]. Liu et al. [88] extracted two major static task facets, *task product* and *task goal*, from Li and Belkin [78]'s faceted framework and used the combination of these two facets to represent and measure task complexity. They found that tasks of different levels of complexity (e.g., factual-product, specific-goal tasks, intellectual-product, amorphous-goal tasks) can be represented by different patterns of local information seeking intentions in search iterations and search actions. Sarkar et al. [116] also represented task complexity using the unique combinations of task product and task goal and investigated the patterns of the search problems that users encountered during task process. They found that in tasks of varying levels of complexity, users encountered different search obstacles, preferred different supports from systems, and adopted distinct search strategies. Moreover, this difference in task complexity, encountered problems and preferred system help can to some extents be inferred from users' search behaviors.

The last few years have seen the information seeking and IR communities tackle more complex search tasks that involve multiple rounds of distinct search actions and active transitions of cognitive states [35, 71]. Many of the existing studies represented different levels of task complexity using one (e.g., task determinability, complexity of learning goals) or more (e.g., combinations of task facet values) static task features and revealed some of the behavioral effects of complex search tasks. However, there has been little data-driven work representing complex tasks as sequences of cognitive and behavioral 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 distribution and dynamic transitions of multidimensional task states. This analytical approach is critical especially for developing adaptive search systems that can go beyond simple, static task properties and support users according to their current task states (i.e. reactive support [59]) or even the prediction of their subsequent states (i.e. proactive supports [90]).

2.2 Information seeking intentions

Search task is a complex, multidimensional concept. To decompose search task, prior studies mainly focused on a set of widely-discussed search task features (e.g., task complexity, task difficulty, task type) and their connections to the variations in Web search behavior (e.g., [5, 69]). Regarding the cognitive space and intention aspect of users, however, information seeking intentions in query segments as cognitive-level components of search task have been scarcely studied. As a result, although users often navigate to useful information with small, local steps under the influence of search tasks [131], it is still unclear how their information seeking intentions in local steps (i.e., query segments) are connected to low-level search behavior and high-level, global context (i.e., search task).

Some early theoretical research on classifying search sessions could be thought of as applicable to information seeking intentions. For instance, Broder [21] argued that Web searches in terms of information seeking and search intention could be classified into three categories: navigational, transactional, and informational. Aiming to build a more detailed scheme of intentions, Kellar, Watters, and Shepherd [68] proposed a new typology of user intentions generated in search, which includes fact finding, information gathering, browsing, and transactions. These classifications, as is some later work discussed below, were built upon analysis of queries that initiated search sessions and did not take into account the search actions within query segments. Also, these early theoretical classifications only identified broad intention categories and ignored the nuance between different specific intentions within each category.

Besides the theoretical speculation on the classification of intentions, according to Rha et al. [108], Xie [150] is the only example of an empirically-based classification of intentions which motivate people to engage in different interactions with search engines. Nevertheless, other similar research on users' goals, knowledge gap, and search intent can also be of help to understand users' intention in information seeking and search episodes. For example, Rose and Levinson [112] analyzed a set of queries randomly selected from AltaVista query logs and proposed a hierarchical typology of users' search goals. Similarly, drawing on the ideas of the Sense-Making approach [36], Savolainen and Kari [120] revealed the discontinuous and dynamic nature of Web searching episodes and developed a conceptual framework of knowledge gaps faced by searchers as well as the corresponding gap-bridging strategies. Jansen and Booth [62] developed a three-level hierarchy of user intent to automatically classify Web search queries based on the information seeking intentions behind these queries. They found that users' query intent (i.e., informational, navigational, transactional) varies by different search topics. In recent research on information seeking intention, Cole et al. [34] studied users' page use and eye movement patterns in search tasks of different types and extracted a set of behavioral signals which may help detect the transitions of users' cognitive states and intentions in search iterations. Mitsui, Shah, and Belkin [101] developed a set of information seeking intentions based on Xie [150]'s initial typology of interactive intentions and empirically investigated the distributions of different intentions in search tasks of different types. Rha and her colleagues studied how different types and states (i.e., satisfied or unsatisfied) of information seeking intentions lead to different query reformulation reasons and strategies [108, 109]. To develop a more comprehensive, multi-level understanding of Web search, Liu et al. [88] investigated and explained the connections between static search task features (i.e. task product and task goal), the distribution and transitions of information seeking intentions, and Web search behaviors.

If we consider the totality of all intentions as users' cognitive space within a task environment, then the specific intentions are essentially separate dimensions of the cognitive space, rather than different values of the same categorical variable. In this sense, the combination of individual information seeking intentions can be considered as a valid representation of users' task states at cognitive level. Hence, mathematically, each cognitive task state can be operationalized as a vector which includes the values of different intentions (i.e. intention present or absent) as separate elements.

2.3 Problems and barriers in search interactions

Individuals face various difficulties and search barriers in tasks of different types, which often cause uncertainties, reduced level of user satisfaction, and even search frustrations and abandonment in the associated information seeking and searching processes [45, 75, 118, 119, 137, 147, 89]. The types of encountered problems and barriers that are explored in information seeking and IIR studies include: internal barriers (e.g., lack of task and topic knowledge, unable to articulate and express information need with the combination of query terms) [15], external barriers (e.g., time constraints, institutional restrictions, information overload, issues in search results presentation) [117, 137, 31, 125, 83], and interpersonal barriers (e.g., lack of effective help from other people) [130]. Within Dervin's Sense-Making framework [36, 158], the specific problems and barriers encountered during search can be considered as the representations of the gap between a user's current situation (e.g, state of knowledge, skills possessed, supports and constraints) and the desired outcome (e.g., work task completion, information need(s) being satisfied).

When users encounter problems and barriers in task-based search processes, they often seek for assistance either from interactive intelligent systems (including IR systems) or people (e.g., friends and families, experts on specific topics) to resolve the problems [61]. Xie and Cool [152] investigated the tasks and problematic situations that often arise in digital libraries as well as the associated help-seeking behaviors. They identified 15 distinct types of help-seeking problematic situations that lead novice digital library users to seek for help. Shah [121] argues that based on the nature of encountered or predicted problems, an integrated interactive recommender system should provide not only the traditional types of system supports (e.g., query suggestions and auto-completion, search results recommendation), but also other potentially useful recommendations, such as people recommendation.

In sum, when seeking and searching for information, individuals usually encounter problems (especially in complex, difficult tasks) and thereby seek for useful system supports. From the process-oriented perspective, this phenomenon encourages us to explore the nature of complex search tasks on a new, situational dimension: what are the distributions and transition patterns of the search problems encountered and user-preferred help in search tasks of different types? How can an interactive search system leverages the knowledge of in-situ search barriers and generate dynamic supports for users in real time? More detailed discussions on the relevant personalization and search recommendation models are included in the following sections.

2.4 Search behaviors

People often search for information on the Web to resolve the encountered gaps in sense making, tasks, or information supply-demand disequilibrium [86]. Several types and components of search behaviors have already been explored and discussed in the review of studies on the relationship between task facets and search interactions. The levels of search behaviors include: information search strategies, search tactics, and search actions [150, 34, 13, 14]. Information search strategies are usually represented by the sequence of search tactics within a search session. Search tactic as a segment of search strategy consists of the actions within a query segment or search iteration (e.g., query formulation, search result browsing and examination). Action as the most basic component of search interaction refers to the individual “moves” (e.g., formulate a query, examine the first search result, keep browsing the current SERP) at different moments of search iterations. The main components of search behaviors are: query behaviors (i.e. query formulation and reformulation), clicking, page visiting behaviors, and search result examinations and usefulness judgments [88, 30]. Note that these are behaviors that are allowed in current interactive search systems (e.g., commercial search engines), and that there are more behaviors and system supports needed in general information seeking, searching, learning and sense making activities [16, 151, 124, 159]. Supporting general information seeking actions that are not allowed in current search systems is beyond the scope of this dissertation study and will be explored in our future research.

In addition to the search behavioral metrics, some researchers also employed neuro-physiological features (e.g., eye movements, data regarding brain activities from EEG and fMRI) in task feature prediction and user modeling [50, 34, 102, 33, 58]. These features help researchers get closer to the cognitive variations during the search process and examine the relationship between human cognition and search actions.

Many empirical studies have demonstrated that the sequence and values of various search behavioral measures can capture several aspects of search task process and can be applied to disambiguate search tasks over multiple dimensions [84, 34, 69, 39, 53, 55]. However, not many studies have directly explored the underlying connection between temporal variations in other dimensions or aspects (e.g., intention states, encountered search problems) and the associated changes in search tactics and actions. To fully understand the process of a (complex) search task and develop adaptive supports, it is critical to combine multiple types of information (e.g., intentions: what users are trying to accomplish; problems and help: what search problems and barriers are encountered by users at different moments and what types of help they are seeking for; search behavioral patterns) and leverage the information in (1) modeling the nature and transitions of multidimensional task states from process-oriented/state-based perspective) and (2) updating the policy of generating potentially useful system recommendations.

2.5 Personalized search

Personalized search systems seek to go beyond traditional retrieval algorithms and to utilize the knowledge about users when providing search results and recommendations. The supports and outputs from search systems are tailored to each individual user's profile that goes beyond the ad hoc retrieval framework. Micarelli et al. [98] summarizes two traditional approaches to personalizing search results: (1) personalizing through modifying the issued query; (2) updating the weights of terms and ranking algorithms iteratively to re-rank the search results originally retrieved by the search system. Most of the current commercial search engines provide some sort of personalization for users based on the general information obtained from user, such as the location where the query was issued, visited topics, and past search behavioral patterns. Empirical evidences show that search personalization techniques can achieve sizable improvements in search performances and user happiness [133, 155]. These improvements can not be easily reached with updates in core search and recommender algorithms which usually focus on global performance rather than specific individuals and their cognitive states [132, 107, 115].

Personalized search often involves using user profile based information, such as demographic features, geographical location, task properties, inferred search expertise, topical interest, time of formulating the query, as well as other relevant features. Most of these features used in personalization are static, at least within the range of a search session. Little work has been done on dynamically improving session-based search strategy or action at the user level [54, 55]. To address this issue, Hendahewa and

Shah [55] sought to understand user strategies in a dynamic manner based on the actions users take over the duration of exploratory search and evaluated different strategies extracted from different users in order to recommend effective search paths to underperfoming users. Yang et al. [154] employed partially observable Markov decision process (POMDP) in session search modeling, aiming to understand users' transitions among different search states and to optimizes search process through providing personalized, dynamic recommendations. These studies on dynamic modeling and personalization enhance our understanding on how users interact with search systems in complex, exploratory tasks of different types. Our study here seeks to take a step forward by (1) connecting the knowledge of *how* users search (i.e. search strategies, tactics, and actions/moves) with that of *why* users search in such ways (i.e. intention states at different moments; encountered search problems and help needed in the local problematic situations), and (2) leveraging the integrated knowledge about the *how* part and *why* part in representing complex search tasks and generating state-based personalized search recommendations.

2.6 Search recommendations

Given the knowledge about users' task states (e.g., intentions, problems, search actions) on multiple dimensions, what types of system supports and recommendations should an interactive search system offer? To help answer this question, in this section we take a step back and review several relevant empirical studies from the research area of search recommendations.

Recommendation-oriented IR studies focus on offering useful recommendations not only based on the classical ranking algorithms, but also by leveraging the knowledge about user characteristics, developing task models, analyzing the similarity of uses' search trails (e.g., pages visited and rated) and behaviors, item similarities and correlations. In many cases, the suggestions generated by systems are closely related to decision-making processes at different levels, such as what queries to formulate, what items to buy, or what package/portfolio to order. The widely used techniques in building recommendation algorithms and systems can be divided into the following categories [110]: (1) content-based method: learning to recommend items that are similar to the ones that the target user liked or interacted with in his or her own past history) [106]; (2) collaborative filtering: identifying and Recommending the items that other users with similar preferences liked in the past [56]; (3) demographic-based method: making recommendations based on the demographic information collected from the user (especially in addressing cold-start problem) [114]; (4) knowledge-based method: recommending items based on the existing domain knowledge about how certain properties of the items can fulfill users' needs [22]; and (5) community-based method: recommending items based on the preferences of the target user's certain group(s) of friends [67].

In information seeking and IIR, developing a useful, stable task or user profile is a fairly challenging task as the task and user profiles extracted from search iterations would change throughout the task process. Hence, constructing and evaluating a task state model in a dynamic manner may be helpful for reconciling the conflicts between static task or user models and dynamic search interactions [55]. Current recommendation systems in information search and retrieval practices are mostly about query auto-completion and recommending related search topics and terms. However, given the varying levels of tasks and the diversity of information seeking intentions and search barriers, users may need system help and recommendations of different types at different moments or states of tasks, such as SERP recommendation and search path recommendation [54, 55]. Also, due to the constant changes in local search contexts, it would be difficult for researchers to evaluate search systems during a complete complex search process with the same set of evaluation metrics. For instance, if an intention at some point of search is to find a potentially useful information object, then precision or nDCG (normalized discounted cumulative gain) [63] is an obvious candidate. However, if the current information seeking intention is to learn domain knowledge or explore a topic, then we need to measure and evaluate the coverage and quality of information retrieved in the related query segment. The selection of specific evaluation metrics determines the way in which we evaluate the performance of interactive search system and define the rewards obtained from taking certain search recommendations.

2.7 Markov decision process

Given the dynamic nature of task-based search interactions, IIR researchers have explored several ways to characterize the sequence and transitions of users' search phases and tactics. One widely-used approach is to simplify the process by making it memoryless (assuming that the current state is dependent only on its previous state) (e.g. [38, 134, 134, 94]). In the IIR context, this assumption is reasonable in the sense that users' often decide their search tactics in local, small steps (i.e. query segments), rather than consciously making and executing global search plans (e.g., search tasks and work tasks) [131, 151]. This is known as the Markov Property [57]. A Markov Decision Process (MDP) is a stochastic decision process that has the Markov property.

Formally, a MDP consists of five components [18]: (1) states: the status the agent is in at a given moment during the interaction process; (2) action: the possible moves or actions that the agent can take at a given moment; (3) reward: the benefit of taking a specific action in a particular state; (4) state transition function: the probability of transition from one state to another state triggered by a specific action or move; (5) discount rate: discount factor for future rewards estimation. The ultimate goal of an MDP is to find an optimal policy that identifies the best actions (i.e. obtaining maximized rewards) for a sequence of states. Q-learning as a model-free reinforcement learning algorithm can automatically learn a policy from the processes and outcomes (e.g., reward, state transition) from the interaction between an agent (e.g. searcher) and the environment [138]. The Q-learning algorithm can keep updating and improving current policy based on the accumulated partial information about the environment and eventually converge to an optimal policy which maximizes the expected value of the total reward or benefit over all successive steps of interaction episodes [97].

In IR research, Luo et al. [94] applied partially observable MDP framework in characterizing and optimizing task-based search processes. Specifically, they proposed four hidden decision making states based on users' query term selection and page visiting behaviors. The four states were defined by two dimensions: (1) relevance dimension: whether the user thinks the returned documents are relevant. Luo et al. [94] argued that if the set of previously retrieved documents leads to at least one SAT clicks (dwell time longer than 30s on the clicked pages, then the current state is likely to be relevant. (2) exploration dimension: whether the user would like to explore another subtopic or keep searching within the current information patch. If the newly added query term(s) appear in previously retrieved documents, then it means the user stays at the same sub information need and thereby is likely to continue exploitation. Within the reinforcement learning framework, Luo et al. [94] mathematically modeled dynamics in task-based session search and simulated optimized recommendations based on the estimated states in the interaction process.

Luo et al. [94] offers a starting point for applying MDP-based reinforcement learning approach in IR research problems and demonstrates the potential of this approach in modeling search sessions and simulating dynamic recommendations based on partial information about the context of search interaction. One limitation of their work is that we do not know what exactly users were trying to accomplish at different moments and what problems and barriers they encountered when trying to satisfy their local information seeking intentions. Therefore, we still need a dynamic state-based framework which can (1) clarify the connection between how people search and *why* people search in such ways and then (2) leverage the knowledge about the *how* and why parts in developing state-based personalized search supports and recommendations. We believe that this is a good place where interactive IR insights can come into play.

2.8 Summary: Theoretical framework and research questions

This dissertation study seeks to capture the dynamic aspect of complex search tasks and represent task states based on three dimensions: (1) intention state (active dimension): the meaningful combinations of individual information seeking intentions in query segments; (2) problem-help state (situational, unplanned dimension): the problems users encounter in search interaction and the types of system help needed; (3) search behavior dimension: the search tactics in query segments.

Figure 2.6 illustrates a broad theoretical model where the dissertation study is situated in. In this model, based on the facets identified in [78] and other relevant works (e.g, [69]), we separate the static task facets and dynamic task facets. The static task facets mainly include predefined task goal, task product, cognitive task complexity, user characteristics (e.g., long-term memory and knowledge base), and so on. The dynamic task facets refers to the facets that vary across different states of tasks, such as perceived task difficulty, users' expectation and satisficing level regarding search performance, knowledge about the specific task topic. Recognizing the difference between these two groups of task facets can help reconcile the conflicts between static task properties and the dynamic nature of task-based searching at both theoretical and empirical levels.

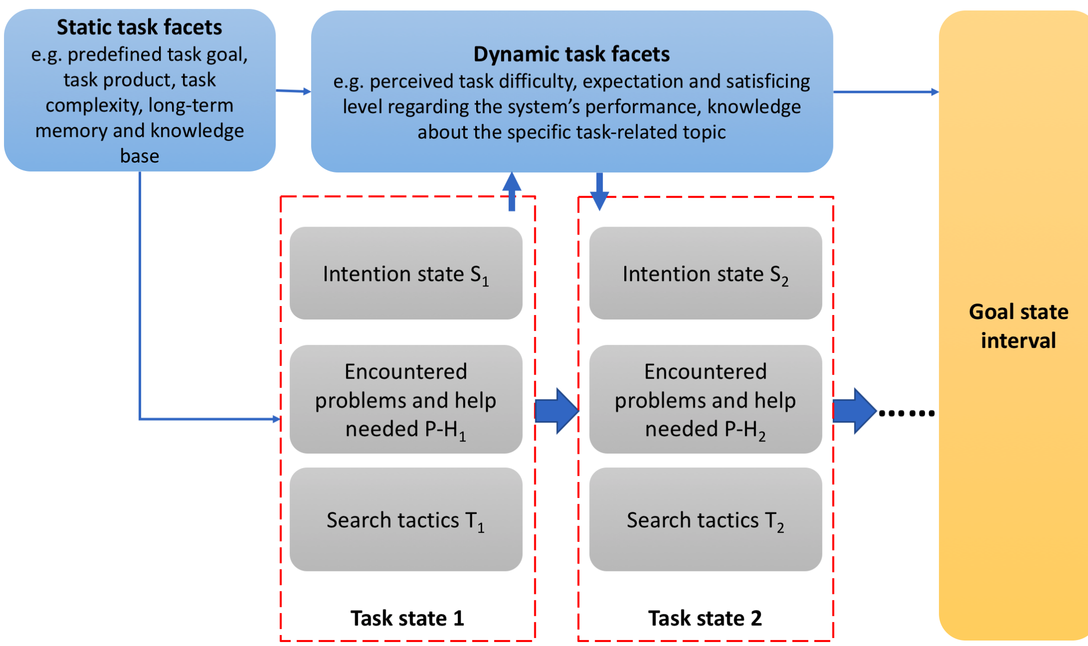


Figure 2.6: Theoretical model of task-based search session.

In addition, we represent the final state or outcome of a search task as a *goal state interval*, instead of a fixed, desired outcome (represented by a single point). This is because the goal state of a search task in users' minds often change over time during the search process [104, 151, 131]. For instance, after some failed attempts in searching for some specific items, the user may consider these items as not available in the search system and thereby remove them from the goal(s) of the search task or change the emphasis and strategy of search. Also, early successes in information searching may encourage the user to expand the predefined search goal and explore broader relevant topics and areas of interest. Therefore, the presumably specific goal state or point of search task may keep moving within the interval during the search process until the final position is determined by the balance among users' search skills, information gains, and the informational requirements for addressing current work task and the underlying problematic situation. The boundary or endpoints of a goal state interval is likely to be decided by both the overarching work task and the ongoing search process (e.g., information gains, problems and failures encountered so far, changes in knowledge state). This dissertation work does not involve the investigation of goal state interval. Instead, we mainly focus on studying the nature and transitions of task states.

Based on the research gaps identified in the literature review above, we seek to answer the following four research questions:

- **RQ1**: What are the states of complex search tasks?
- **RQ2**: What are the transition probabilities between the states in search tasks of different types?
- **RQ3**: To what extent can we infer and predict task states from search behavior?
- **RQ4**: To what extent can we improve users' search efficiency by offering state-based search supports in complex search tasks of different types?

The first two research questions speak to the goal of understanding dynamic task states and state transition patterns. This is an initial step towards conceptualizing complex search tasks from a process-oriented perspective. The answers to these two questions can help us advance knowledge in the first aspect: representing and explaning the dynamic aspect of complex search tasks using the distributions and transition patterns of dynamic task states. Taking a step forward, RQ3 corresponds to the second main question proposed in the Introduction chapter and addresses the connection between observable search behaviors and implicit task states in search interactions. With respect to practical application, RQ4 explores the possibility of leveraging the knowledge of users' task states in improving search efficiency and adaptively supporting users engaging in complex search tasks.

Chapter 3

Methodology

Major constituents of the proposed dissertation work include: (1) identifying task states (RQ1) and investigating the transitions of task states in complex search tasks of different types (RQ2); (2) exploring the relationships between two impactful dimensions of task states (i.e. intention state and problem-help state) and various aspects of users' search behaviors (RQ3); (3) simulating state-based search recommendations and evaluating them based on task-oriented reward metrics (RQ4). With the knowledge learned from rich, small-scale user study datasets in this dissertation study, the ultimate goal of this line of research is to develop a scalable, non-intrusive approach that can iteratively evaluate search interactions at different states in an online fashion and provide necessary recommendations and interventions without hindering the ongoing search interaction process or relying on explicit in-situ feedbacks and judgments from users. In this dissertation study, recommendations refers to *the search paths that are recommended by the adaptive search system (simulation algorithm) under different task states.*

This chapter explains the concepts, data collection techniques, analytical models and evaluation measures associated with each research question. In the flow of this dissertation research, we start with investigating different aspects of users' task states and then analyze state transition patterns under tasks of different types. Furthermore, to explore the connections between different levels of search interaction, we examine the extent to which we can infer and predict task states from search behavioral measures using several Machine Learning classifiers. Built upon the knowledge we learned from the analyses above, in the final step we focus on the practical value of our state-based approach and simulate adaptive search path or query segment recommendations using Q-learning algorithm. The overarching goal of this dissertation is to build a state-based framework that can help us deconstruct tasks and support users engaging in complex search tasks from a dynamic, process-oriented perspective.

As a preparation for building analytical models and answering the four proposed RQs, this dissertation work employed the empirical data collected through two user studies (Lab study 1: Information seeking intentions, ISI; Lab study 2: Problems and help, P-H). The ISI study was designed and conducted by a research team the dissertation author belongs to and resulted in a series of publications on users' information seeking intentions, tasks, and variations in search strategies [34, 100, 88]. The ISI study was sponsored by National Science Foundation (NSF) Award #1423239. The P-H study was designed and conducted by the dissertation author and his colleagues and was sponsored by was sponsored by National Science Foundation (NSF) Award #1717488. The early findings from the P-H study enhance our understanding of users' encountered search problems at different points of search sessions and shed light on the connection between in-situ search problems and multiple aspects of Web search behavior [116]. This dissertation study combined the data from these two studies focusing on two different aspects of complex search tasks and aims to capture task state transition patterns and leverage the knowledge about task states in supporting users' search activities.

3.1 Lab study 1: Information seeking intentions

The information seeking intentions (ISI) study gathered intention annotation and search behavior data from 693 query segments generated by 40 participants in tasks of four different types. The average session length is 8.66 queries (Min.= 2 queries, Max.= 29 queries), which offers relatively rich data for dynamic modeling and personalization. The study techniques and procedure are explained in detail in the following sections.

3.1.1 Intention taxonomy

To classify users' information seeking intentions in Web search query segments, the Lab study 1 used the typology of search intentions which was developed and elaborated by

Rha et al. [108] based on a subset of Xie [150]'s classification of interaction intentions. Participants were given a description of this typology to the participants before their search sessions were replayed for intention annotation. Then, participants were asked to identify their information seeking intention(s) for *each query segment* based on the typology. Participants could identify multiple intentions in the cases where they tried to accomplish multiple things within a single query segment. Therefore, the intentions identified in this study are not mutually exclusive and should be considered as different variables. Details regarding the intention annotation process are provided in "Procedures" section below. The twenty information seeking intentions are listed in Table 3.1.

Table 3.1: Information seeking intentions and the associated acronyms

Intention Type	**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)

3.1.2 Search task and participants

Four work tasks within the domain of journalism, originally designed by Cole et al. [34], were employed in the ISI study for controlling the possible task effect: Copy editing (CPE), Story pitch (STP), Relationships (REL), and Interview preparation (INT). Forty undergraduate students majoring in journalism were recruited as participants from a research university. To ensure that the participants have a certain level of familiarity with journalism tasks, only upper-division undergraduates who have completed either one journalism writing or reporting class were selected.

The four different search tasks types were defined using four task facets extracted from the task classification scheme proposed by Li and Belkin [78], as modified by Cole et al. [34]: Product, Goal, Level, and Named. With respect to task Product, intellectual task refers to a task which produces new ideas or findings, whereas factual information task refers to a task locating facts, data, or other similar information objects in IR systems. Regarding task Goal, task with specific goal refers to a task with a goal that is explicit and measurable. By contrast, task with amorphous goal is defined as a task with a goal that has no explicitly defined outcome which cannot be measured in a quantitative sense. In terms of Level, a task can be classified into two categories: 1) document-level task: a task for which a document as a whole is judged; 2) segment-level task: a task for which a part or parts of document are judged. The Named facet refers to whether what is to be found is explicitly named; its values are true or false, named or not named. The values of each value for each of the task types are shown in Table 3.2.

Each of the four task types has two specific versions, which differ from one another in topic (i.e., coelacanths; methane clathrates and global warming). These two topics were selected to control the variable of participant familiarity with the task topic, as our participant population was thought likely to not be familiar with either. Table 3.2 shows the four specific tasks for the topic of coelacanths; "methane clathrates and global warming" was substituted for "coelacanth" to generate four more specific tasks. Each of the participants was asked to do two tasks of different topics and types, in Latin Square design, pairing CPE with INT, and STP with REL, to balance tasks by topic, and by the values of Product, Level and Named facets.

3.1.3 Procedure

Figure 3.1 illustrates the procedure of ISI controlled lab study. The study started with a demographic questionnaire and a tutorial video on the additional browser interface features our plug-in provided for participants in information search sessions. Participants were free to search anywhere on the web, with the only restriction being to conduct their interactions within the supplied browser, which logged their search behaviors, and provided the ability to save and unsave web pages. The participants then went through

Table 3.2: Tasks assigned in lab study

Task Type	Task Descriptions
Copy Editing (Factual Specific Segment Named True)	**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 Segment Named False)	**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 Document Named True)	**Assignment**: You are writing an article about coelacanths and conservation efforts. You have found an interesting article about coelacanths but in order to develop your article 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 Preparation (Intellectual Amorphous Document Named False)	**Assignment**: 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 potential impact of Dr. Erdmann's work.

the task description (see Table 3.2) and answered a short questionnaire on task familiarity/experience, topic familiarity, and anticipated task difficulty. They then had up to 20 minutes to complete the first assigned search task, by searching anywhere on the Web, with any search facilities that they wished, but could choose to finish early if they completed it to the best of their ability. Participants' search activities (e.g., query, timestamp, URLs, retrieved pages and documents, page type) were recorded with a Firefox browser plugin and Morae (https://www.techsmith.de/morae.html).

Afterwards, participants were asked to finish a post-search questionnaire on search task difficulty and self-reported search experience. At this point, they were asked to read a guidance of the intention annotation task and to watch a video explicating how to annotate search intention in each query segment. We then replayed the entire search, query segment by query segment, asking for intention annotation for each segment, in sequence (see Figure 3.2). In our study the intention annotation task had no time limit. For the intention annotation task, participants were asked to select which intentions applied to each query segment in the search session. Within each query segment, participants could choose any number of intentions from the list. Participants could

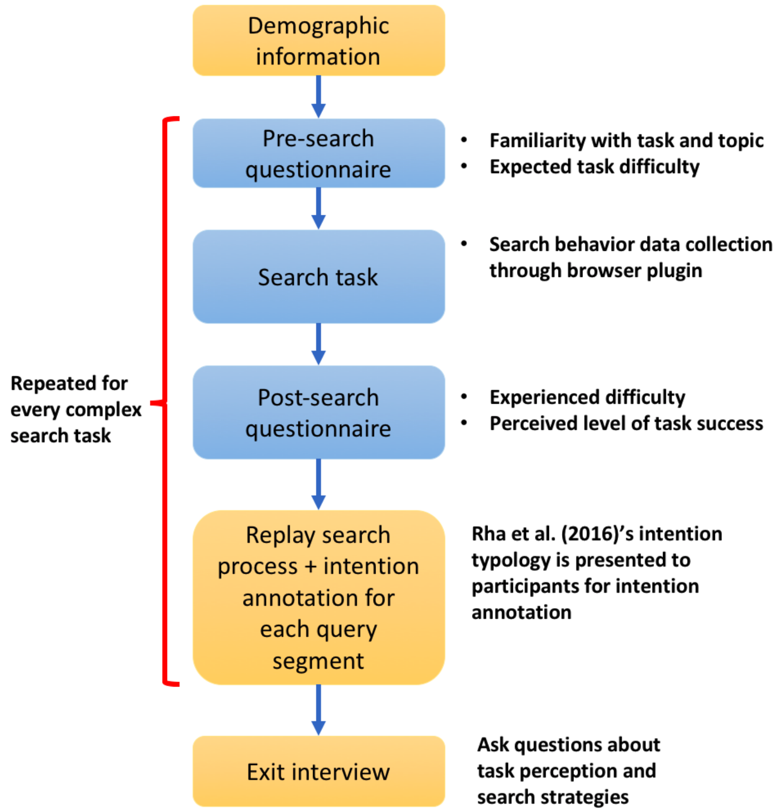


Figure 3.1: ISI study procedure.

choose "other" if their intention did not match the 20 intentions provided. Participants repeated this selection process for each query segment during the course of intention annotation. The same procedure was then followed for the second search task, and the study session ended with an exit interview with questions related to participants' experience of performing the two search tasks. The entire process took about two hours for each participant.

3.2 Lab study 2: Problems and help

The problem and help (P-H) study focused on the problems that users encountered at different moments of search and the types of system help they needed for addressing the problems. In our controlled lab study, each participant performed three simulated search tasks (one warm-up task and two formal tasks). We collected 273 query segments from twenty-six participants, along with their annotations on encountered problems and preferred help. The user study procedure is explained in detail as follows.

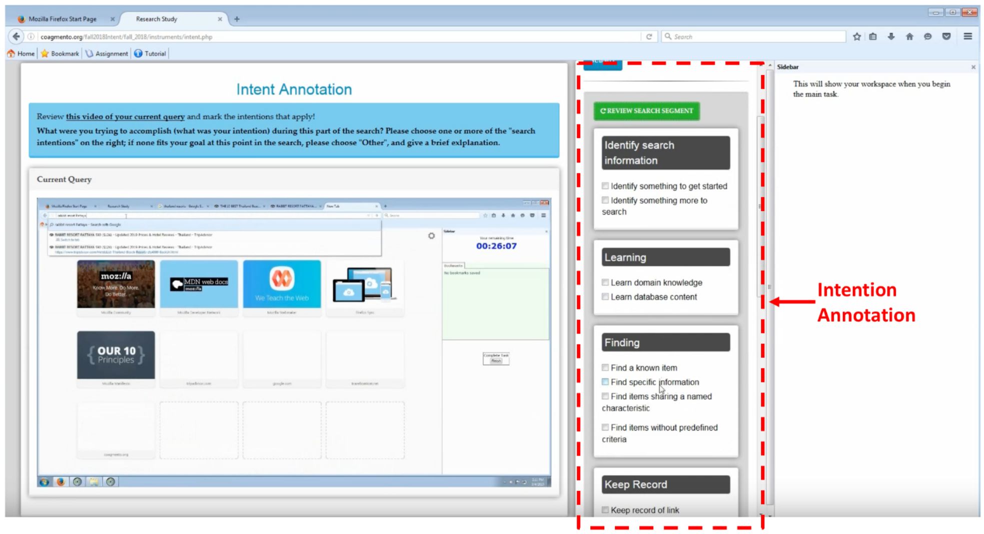


Figure 3.2: Problem-help annotation interface.

3.2.1 Participants

Twenty-six undergraduate students were recruited as study participants at a U.S. research university via various channels, including advertisements to emails lists, Facebook groups and in-class recruitment. 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 Business Administration. 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.

3.2.2 Study procedure

Participants conducted their web search using Google Chrome in a desktop computer equipped with necessary applications in our lab. 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 asked to search for information that would be potentially useful for completing the assigned tasks. Each time they opened a new tab to begin formulating/reformulating a query, the Chrome extension displayed the aforementioned problem/help questionnaire, aiming to collect information about users' in-situ problems and preferred help. Also, to learn about participants' task expectations and perceptions, before and after each search task, we asked participants to answer 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 supports, perceptions of their search experiences in everyday lives, as well as their visions of an ideal information retrieval system performance. The length of a laboratory session for each participant ranged from 50 to 80 minutes total. Figure 3.3 depicts the structure of pre-task, search session, and post-task for a single task.

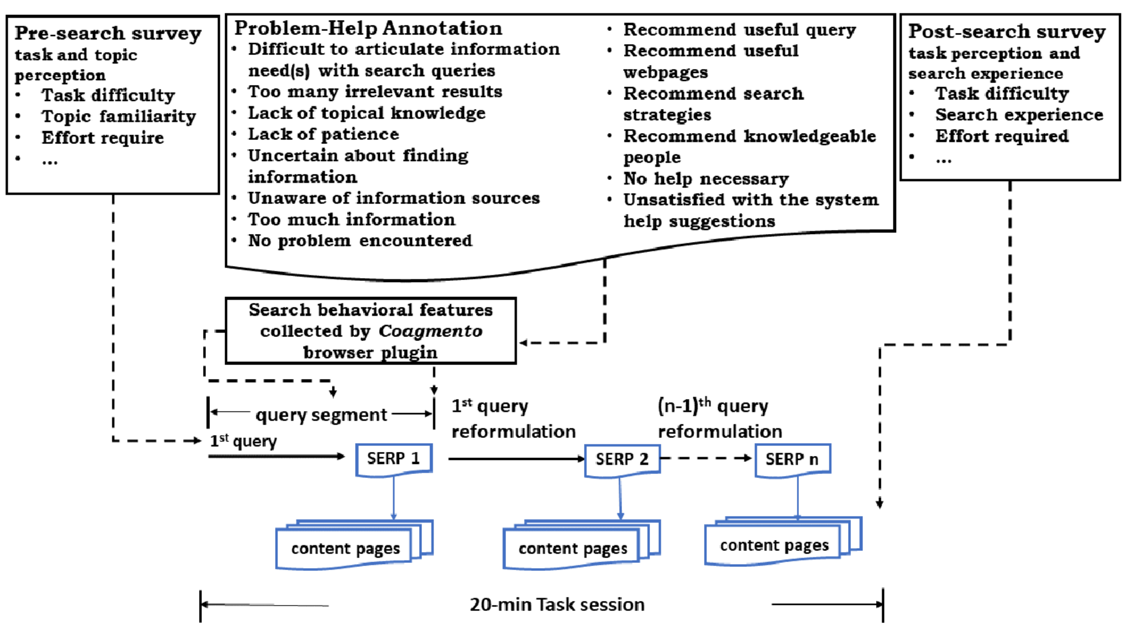


Figure 3.3: Task session flow.

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

3.2.3 User characteristics

Users' own states of knowledge is a major factor that affects users' search tactics and their understanding or estimation of the goal states of tasks at hand. To collect information about participants' knowledge and skills regarding the assigned tasks (i.e., task dimensions, task requirements such as time, and task type), we asked them 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 as well as their state of knowledge after completing the task. Table 3.3 enumerates the questions in the pre- and post-task questionnaires. The questions regarding perceived task difficulty were adapted from [148].

Table 3.3: 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.

3.2.4 Tasks characteristics and their representations

Participants completed three tasks consecutively during their sessions in the laboratory. They started with the 5 minute warm-up task and then two 20 minute formal tasks. Descriptions of each task are provided below. Before each task, participants answered a pre-task questionnaire about their familiarity with the task. In the tasks, participants were asked to bookmark/save useful web pages and to construct brief reports (as responses to the questions in task descriptions) based on their findings from web search. We started the search sessions with a warm-up task, aiming to familiarize participants with the study procedure (especially problem-help annotation and useful page identification), system and laboratory environment.

We designed two complex, multi-round search tasks by manipulating the type of information need, task goal, and task product based on the task classification scheme proposed by Li and Belkin [78]. Their faceted framework identified various characteristics of a task and many of them 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). The three binary labels (i.e. type of information need, task product, task goal) 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 (based on the simulated work tasks of [20]). The task descriptions provided to participants are as follows:

Warm-up Task [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 [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 [Need = Social; Product = Intellectual; Goal = Amorphous] (20 min): *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 (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.*

Table 3.4: 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.

3.2.5 Encountered search barriers and help needed

As mentioned in the previous section, the primary unit of analysis in this work is the perceived problems and help at various points in search sessions. To elicit the information from participants, we provided a questionnaire to them at the beginning

Table 3.5: 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.

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 displayed a pop-up window with a problem-help questionnaire waiting to be answered (see Figure 3.4). Participants were asked to report the problems they just encountered and the possible supports they would like to have in real-time search scenario on the problem-help questionnaire. By virtue of this design, it was possible to collect information about the in-situ 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 help from users, we extracted relevant questions and constructs from existing research on gaps, barriers and help in information seeking and sense-making process (e.g. [137, 118, 31, 152]). The problem and help questions presented in the pop-up window are listed in Tables 3.4-3.5. We converged and combined different types of help, barriers, and problems identified in the aforementioned studies and defined eight types of problem and four types of help which may 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 help 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 make sense in these situations. To avoid potential ordering effects, the order of options for all questions presented in the pop-up windows was randomized.

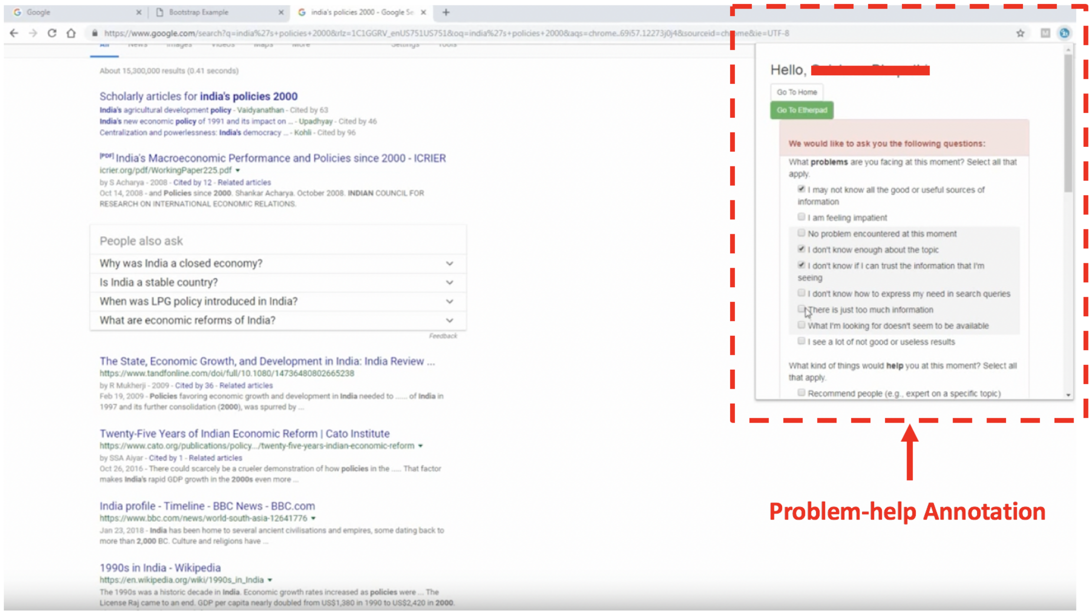


Figure 3.4: Intention annotation interface.

3.3 States and state transition probabilities

Built upon the datasets collected from the two user studies introduced above, this section reports the concepts and methods defined for answering the following two RQs:

- **RQ1**: What are the states of complex search tasks?
- **RQ2**: What are the transition probabilities between the states in search tasks of different types?

In this work, to embrace a balanced understanding of task-based search interactions, we investigate both active and passive, unplanned sides of search task states. On the *active, intentional* dimension, we focus on the *intention states* in query segments or each search iterations. Intention state refers to the thing(s) that a user seeks to accomplish within a query segment. On the *situational, unplanned* dimension, we explore the *problem(s)* encountered by users at query-segment level during the search process and their preferred *help* for tackling the problem(s).

Different intention states represent different combinations among individual intentions. To operationalize the concept of intention states in our work, we do cluster analysis based on users' intention annotations on twenty intention dimensions, aiming to obtain a fixed number of separate clusters as intention states. With respect to the situational, unanticipated dimension, we adopt participants' annotations on problems and preferred help at different moments of search and use all distinct combinations of problems to represent problem-help states.

RQ1 and RQ2 speak to the description of states and state transitions in complex tasks. To address these two research questions, we conduct *clustering analyses* based on intention dataset (from Lab study 1: information seeking intentions) and P-H dataset (from Lab study 2: Problems and help) respectively and seek to extract meaningful clusters as task states on both active and unanticipated dimensions. Then, based on the state transition data from both lab studies, we compute the transition probabilities between different states. The motivation and rationale behind these analyses is that from a process-oriented perspective, complex tasks of different types can be defined and conceptualized as divergent sequences of states and state-transition probabilities. This process-oriented perspective can help us better understand the dynamic nature of complex search tasks.

Specifically, for state identification, we use *K-modes clustering analysis* for extracting clusters out of the user annotation data. K-modes clustering as a unsupervised learning method extends the traditional K-means paradigm to cluster categorical data by using (1) a matching dissimilarity measure for categorical items, (2) modes instead of means for forming clusters, and (3) a frequency-based approach to iteratively update modes in the clustering process to minimize the cost function (which involves assigning a data point to the nearest possible cluster) [29]. In clustering analysis, we represent each query segment using a vector consisting of a variety of elements (corresponding to different information seeking intentions or problem-help items) and group query segments into different categories based on the distance from each query segment (vector) to the *centroid vector* of the cluster. Differing from K-means clustering, in K-modes clustering the distance is measured based on the number of elements where the two vectors do not share the same value. The centroid vector is the mode that minimizes the distances or dissimilarities between the vector itself and each object of the data within the cluster. Thus, unlike K-means clustering, the distance function here is frequency-based. The iterative clustering stops at the point where the total distance of clustering does not decrease. In the intention study (ISI) dataset, task states are represented by the clusters of intention vectors (each of the vectors consists of twenty individual intentions as separate elements). Similarly, in the P-H dataset, task states are represented by the clusters of problem-help vectors (each of the vectors consists of eight different search problems and six help types extracted from previous literature and our pilot studies). It worth noting that clustering algorithm itself does not generate meaningful state type or label. After the valid clusters are generated, we interpreted each cluster and define labels or names of task states based on the most frequent information seeking intention(s) or problem-help combinations in the corresponding clusters.

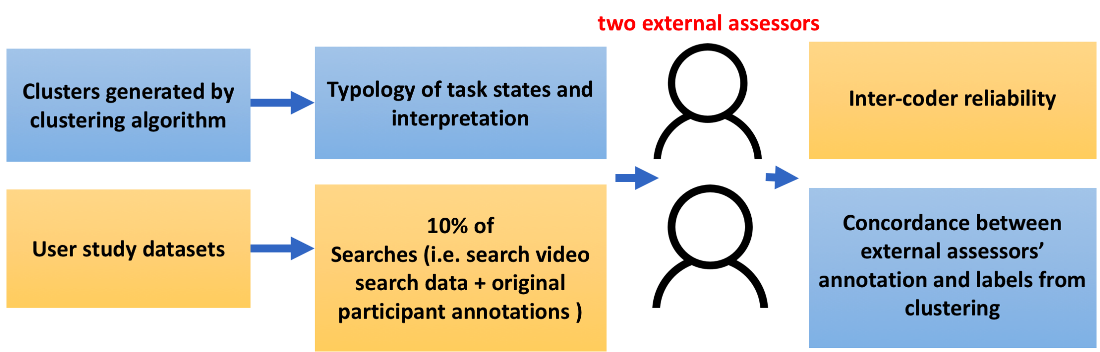


Figure 3.5: External judgment of task state labels generated by the clustering algorithm.

To test the validity of the task state categories extracted from annotation data as well as the labels we defined, we ran *external judgment of state types* with two external assessors. Specifically, we randomly extracted 10% of searches from each type of task and asked the two assessors to manually annotate task state for each query segment independently according to the task states we extracted and defined. Each assessor was provided with the video of participants' search process, the intention or problem-help annotation and search behavior data, as well as the state typology. When annotating the task state of a given query segment, each assessor started with watching of the original video of the corresponding query segment recorded by Morae software and making notes about participants' search actions and tactics (e.g., mouse positions and movements, clicks, text selections, and page examinations). Based on the notes, the assessor had an initial assignment of the state label (selected from the task state labels determined based on the results of clustering) for the query segment. Then, the assessor went through the associated search log data (including query issued, pages visited and dwell time on each page, SERP dwell time, bookmarked pages) and original annotations of information seeking intentions, problems and help and adjusted their state annotation accordingly if the assessor believed that it was necessary to do so. This annotation process was repeated by both assessors for every query segment included in the extracted 10% of searches.

To measure the validity of task state labels, we computed three *Cohen's Kappa coefficients* κ, 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 assessors here. These two assessors as IR researchers are experienced in cleaning and analyzing various type of data generated through search logs, interviews, and recorded videos. The *Cohen's Kappa coefficient* κ can be computed using the following equation:

$$\kappa = \frac{p^0 - p^e}{1 - p^e} \tag{3.1}$$

where p^0 refers to the observed agreement among annotators (i.e. accuracy), and p^e represents the hypothetical probability of random agreement. *Cohen's Kappa coefficient* κ is generally considered to be a more robust reliability measure compared to simple percentage agreement or accuracy as it takes into account the possibility of the agreement between raters occurring at random [76]. After we identified and validated task state labels, we incorporated them into *Markov Chain models* and 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 (RQ2).

3.4 Task state prediction

To answer RQ3 (To what extent can we infer and predict task states from search behavior?), we investigate the implicit connection between users' search behaviors and the associated task states extracted from previous steps. Specifically, we (1) test the correlations between task states and various behavioral measures; (2) examine the extent to which we can infer and predict task states using search behavioral features with several widely used supervised learning (classification) algorithms, such as support vector machine (SVM), random forest, XGBoost, and logistic regression. The performance of prediction analysis is measured based on the overall accuracy of predicting each task state. For inferring and predicting task states, we use all behavioral data up to the current state within the associated search session. The prediction models utilize the following behavioral features, which were extracted from existing IIR evaluation studies and employed in our previous user studies in both lab and field settings.

The query-level search behavioral features employed for building classifiers 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 behavioral measures, we build three types of feature groups 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 session (before current query segment).

To evaluate the prediction performance of our classifiers in a more solid manner and reduce potential random errors, we randomly slice data with an 80/20 split for training/testing in each round and run 500 rounds of training and prediction. Then, we gather all results from prediction models and statistically test the difference between the performance of the selected classifiers and that of the best-performing baseline model.

3.5 State-aware, adaptive search path recommendation

With respect to practical value of our process-oriented approach, based on the knowledge learned from answering the above questions, we apply the proposed state-based framework in supporting users engaging in complex search tasks. To answer the RQ4 and build a demonstration of our approach, we employ Q-learning algorithm and simulate search path or query segment recommendations based on users' task states and the estimated task-oriented rewards (e.g., collecting useful information or pages for completing the task at hand). Then, we evaluate our model by comparing the performance of simulated search path with that of the average performance of participants' original search sessions and measuring the extent to which the simulated recommendations can reduce the efforts or steps of completing assigned tasks.

In the following subsections, we first introduce the key components of our model: states, actions, rewards and evaluation measures. Then, given the defined components, we explain how our simulation model operates on our user study dataset and produces state-based search path recommendations.

3.5.1 States and actions

A Q-learning algorithm is trained to decide actions based on the current state as well as value of state-action pair [129]. In this dissertation, we employ the task states identified from answering RQ1 as the states in our computational model. In our iterative training and value updating process, the evolving Q-learning algorithm will select and evaluate actions based on the current task state.

Regarding actions, based on the features of Q-learning framework as well as the nature of RQ4, we define actions as *the rules (e.g., distinct ranking algorithms, different combinations of parameters, weights, and values) of selecting a specific search path (i.e. query segment) from a finite set of available options at a given moment of search interaction.* The output of Q-learning algorithm is an improved *policy* that defines the values of all state-action pairs (i.e. the estimated reward a user can obtain through taking a specific action under a given state). This trained policy can determine the optimal action given a user's task state and simulate a recommended search path extracted from a task-based solution space explored by multiple users. The possible actions of our Q-learning algorithm are defined using the following query-level behavioral features which measure different aspects of search interactions:

- Query similarity = (number of overlapped unique terms between two adjacent queries)/(number of total unique terms from the two queries)
- Number of pages clicked within the query segment
- Average dwell time on content page
- Search engine result page (SERP) dwell time

In addition to the four features above, we also take into consideration the content page clicking patterns under different states and tasks and extracted related features using *Word Embedding* technique. Specifically, we employ Word2vec algorithm (in this case, continuous bag-of-words algorithm or CBOW) for producing word embeddings using the unique content pages or URLs clicked in all query segments. We define each URL as a "word" and a query segment consisting of multiple clicked URLs as a "sentence". Using CBOW technique, each URL can be turned into a unique vector which includes multiple elements or dimensions. In Natural Language Processing (NLP) tasks, Word2vec technique can "understand" the semantic differences between words and group words with similar meanings together [99, 111]. In this study, similar or adjacent URLs (e.g., visited content pages that are frequently "co-clicked" or close to each other in multiple query segments) will get high weight values on same dimensions. Using word embedding technique here can help us turn unique URLs or content pages into vectors for training models and may be able to separate highly useful pages from less useful and useless pages. Therefore, we add the trained weights of dimensions from CBOW algorithm to the ranking function and utilize these dimensions to define actions and rank search paths or query segments under different states. Since we are dealing with a relatively small dataset here, we set the number of dimensions as 5 for the hidden layer of CBOW Neural Network (NN) training process.

As it is discussed above, actions of the Q-learning algorithm refer to the rules of ranking query segments under different task states. Inspired by the model setups in previous reinforcement-learning-based IR studies (e.g., [94]), we rank every qualified query segment (i.e. query segment under the same topic and state) based on a linear combination of the weighted ranks of the query segment generated according to the value of different features or dimensions. Given a defined action, the query segment which receives the highest total weighted rank (smallest rank value) will be ranked on the top and thus will be treated as the recommended search path. Based on this definition, we can obtain different actions or rules of ranking by manipulating the weights of one or more features. The Q-learning recommendation process can be described as a "guessing game": At each step or query segment, the algorithm seeks to guess the "best action" that can maximize the chance of obtaining highest rewards.

To generate different actions and broaden the possible action space for training, we begin with the "baseline action" which assigns the same weight to all features and then change the weight of one feature at a time by increasing the weight of the feature by a factor of $x = \{1, 2, 3, 4, 5\}$. At the same time, we evenly decrease the weights of other features. We repeat this process for all features (four behavioral features and five Word2vec dimensions) and generate 37 unique actions. During the state-action value updating and policy learning process, using different actions may lead to different query segment recommendations and thereby produce different rewards for the user. To answer RQ4, we aim to train a Q-learning algorithm or policy that can identify an accessible search path with potentially high reward given the knowledge about a user's task state. To train the Q-learning algorithm and iteratively update the policy behind the model, we need a task-based ground-truth reward measure.

3.5.2 Rewards

Since our ultimate goal is to support users engaging in complex search tasks, we define rewards of query segments based on their actual contributions to completing the task at hand. In our lab studies, we asked participants to search for information and bookmark pages that are useful for completing the tasks we assigned to them. To accurately measure the reward associated with different bookmarked pages and query segments, we annotated bookmarked pages and defined their respective rewards under a certain task context based on the extent to which they helped participants fulfill the requirements stated in our search task descriptions.

Specifically, for example, in the copy editing (CPE) task, we ask participants to search for and bookmark pages that either confirm or disconfirm one of the six statements we provided in task descriptions. Therefore, we annotated all bookmarked pages under this task and represent each bookmarked page with a vector consisting of six elements, corresponding to the six statements in CPE task. If a page confirms the first two of the six task statements, then we represent the page using vector $v = \{1, 1, 0, 0, 0, 0\}$. If another bookmarked page confirms the last statement, then the page can be represented using the vector $v = \{0, 0, 0, 0, 0, 1\}$. Thus, the total reward from these two pages can be measured by the sum vector $v = \{1, 1, 0, 0, 0, 1\}$. In Q-learning, as the simulated search sessions proceed, we add the vectors of all bookmarked pages together and terminate a simulated session when all elements larger than zero (which means all statements are confirmed or disconfirmed by at least one unique statement). Similarly, for the interview preparation (INT) tasks, we apply the same annotation and reward calculation methods and use reward vectors to represent bookmarked pages in training.

Compared to CPE and INT tasks, the story pitch (STP) and relationship (REL) tasks have more subjective and flexible criteria of search task completion. Specifically, the STP task asked participants to bookmark pages that contain six most interesting facts about coelacanths or methane clathrates. In REL task, participants were asked to find an authoritative web page that explains the relationship between two of the five listed facts. For these two tasks, we define the task completion (or simulation termination) point based on users' bookmarking behavior: for the STP task, we terminate a simulated search session once the algorithm collects six unique bookmarked pages within the session. With respect to the REL task, we end a simulated session once the algorithm finds a potentially authoritative content page that was bookmarked by at least two different participants.

To illustrate our reward annotation procedure for CPE and INT tasks, we list four examples of manual annotations in the section 7.1 of the *Appendices* chapter. Through iteratively training our model under a given task context, we intend to identify potentially high-reward query segments which can form an optimized search path and help users complete their search task with less steps or in shorter sessions.

3.5.3 Q-Learning process

Figure 3.6 explains how the Q-learning-based simulation mechanism works in this dissertation study. Note that term "Learning" here refers to the process of exploring the connections between state, action, and task-oriented rewards (i.e. useful web pages) and iteratively updating a policy which can identify the optimal action (i.e. a particular rule for ranking query segments) given a user's current task state.

As a preparation before the simulation actually starts, we need to first define a Q table which facilitates the process of finding optimal policy. The Q table is a table of states by actions that is usually initialized to zero, then each cell (which contains the value or quality of taking an action in a given state) is iteratively updated based on the outcome (i.e. immediate reward and estimate of optimal future value) from the interaction between agent (i.e. user) and the environment (e.g., information presented by the system, task state transitions) [129]. Q table presents the mapping from states to actions and partially[2] determines the selection of action as it shows the expected value of the total reward starting from the current state.

Suppose there is a target user u_t that the interactive search system is trying to help. The target user u_t initializes the search process with a starting task state s_1 (e.g., exploration of the topic; anomalous state of knowledge). Given the knowledge of the user's task state, the search system goes back to the Q function/table, finds the corresponding state, and locates the current best available action a_1 under the given state. The actions are evaluated based on a weighted sum of the expected values of the rewards of all future steps or search iterations starting from the current task state. Note

[2]we say "partially" here because we apply ϵ-greedy algorithm in action selection, which means that there is a certain chance that the Q-learning algorithm will randomly choose an action (exploring the accessible solution space) regardless of the estimated Q values.

that at the initial state we may start the simulation process with a set of arbitrarily defined, fixed values of Q function (e.g., zero) for all state-action pairs. Thus, we may start with randomly selecting an action, which may generate a potentially better search path sp_1 from user u_1.

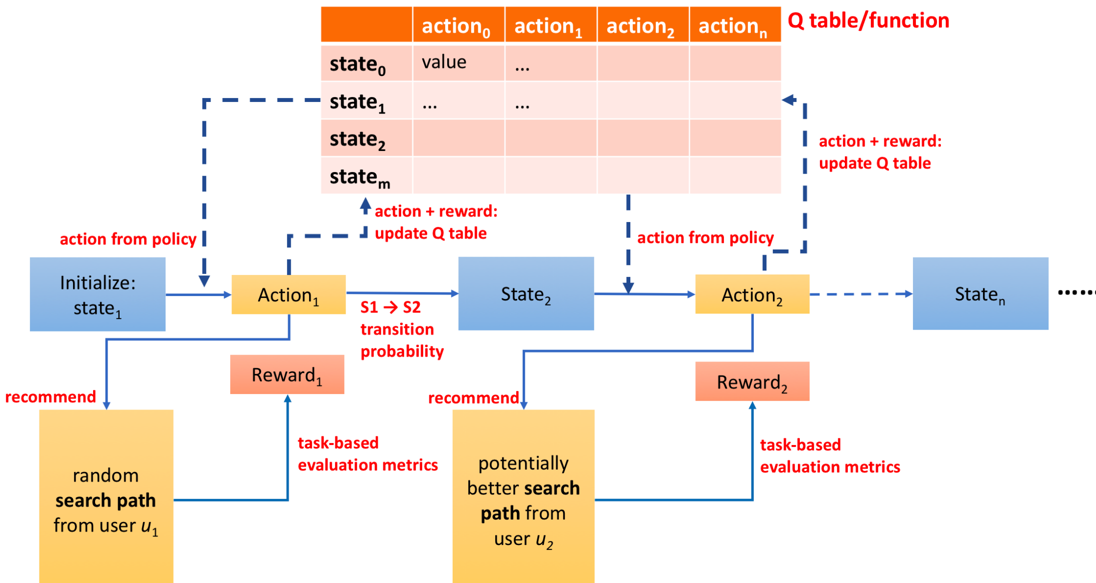


Figure 3.6: State-based search support simulation.

Within the Q-learning framework, besides the immediate reward obtained, the response from search environment also includes the next state of the information seeking and search episode. To keep the iterative simulation process moving forward in a reasonable, realistic way, we add a function $f_{state\_transition}$ which selects the next state s_2 based on the state transition probabilities we extracted from authentic search sessions. The selected action a_1 and the obtained reward r_1 through taking a_1 under the task state s_1 are used to update the values in Q function and improve the current policy.

After that, another round of iteration starts, with the target user u_t moving to a new task state s_2. Similar to the previous round of simulation, the interactive system will go back to the Q function or table again and select an action based on the current new state s_2. The action as a search path ranking algorithm will generate another potentially good search path sp_n from a user u_n. Again, to keep the iterative simulation going, u_n's task state s_n is assumed to be determined by the function $f_{state\_transition}$

based on the corresponding state transition probabilities. The selected action a_n and the received reward r_n through taking a_n under the task state s_n will be taken back to the Q function for updating the values in Q function and further improving the policy (i.e. mapping current state to a potentially better action). We terminate a simulation episode and start another new simulation episode (initialized with a randomly selected state and query segment) once the current accumulated useful information (i.e. annotated rewards from bookmarked pages) satisfies the requirement of search task at hand.

According to the policy improvement process of Q-learning framework [129], in each round of search recommendation simulation, the value function of state-action pairs (i.e. Q function) can be updated as follows:

$$Q^{new}(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha * (r_t + \gamma * max_a Q(s_{t+1}, a) - Q(s_t, a_t)) \tag{3.2}$$

where r_t is the reward the target user obtained when moving from s_t to s_{t+1}, and $\alpha \in (0, 1)$ is the learning rate. $max_a Q(s_{t+1}, a)$ is the estimate of optimal future value within the Markov Decision Process (MDP). The discount factor $\gamma \in (0, 1)$ assigns higher value for the rewards received earlier than those received later.

The methods and techniques introduced above can keep the simulation running and iteratively update the policy of choosing an action a from a finite set of possible actions A. Once the action is decided, the search system can select a potentially better (or even optimal) search path from a collection of search paths. These search paths represent the traces of all users (participants)' collective exploration of the uncertain solution space behind a given complex search task. Within the reinforcement learning framework, our goal here for RQ4 is to find an optimal search episode and evaluate the simulated search episode by comparing it with the target user's original search path within a task-based search session.

However, within the current greedy version of Q-learning framework, it is likely that the algorithm keeps repeating a local optimal action without exploring and testing any alternative, potentially better actions. To address this problem, we will apply a

Algorithm 1: Q-learning process for each iteration.

Data: Set of all query segments M; States S_i associated with each segment; predefined task T; Set of all authentic search sessions under the task T: SE_T

Result: Number of steps for completing the search task T;

Initialization;
Group query segments into different subsets M_{S_i} by state S_i;
foreach *authentic search session* m *in* SE_T **do**
    Take the first state S_0 of the session m as the initial state for simulation;
    while *Task requirement is NOT satisfied* **do**
        Take $action_n$ according to the current state S_i and $Q\{state, action\}values$;
        foreach *query segment* sp_i *in* M_{S_i} **do**
            Calculate the weight of each feature according to the $action_n$;
            Calculate the weighted rank value of sp_i under each feature;
            Calculate the total weighted rank value for sp_i;
        end
        Take the top ranked sp as the simulated search path for the current state;
        Receive the reward r_i (remove the reward from repeated pages);
        Update the policy/Q-table Q according to r_i under S_i and $action_n$;
        Move to the next state according to the state transition function $f_{state\_transition}$;
    end
    Calculate the number of steps used in the simulated session $m^{'}$;
end
Calculate the average number of steps used for all simulated sessions;
Save the updated policy Q for the next iteration of training;
Terminate current iteration;

ϵ-greedy policy here. ϵ-greedy policy allows the Q-learning algorithm to (1) select an action with 1-ϵ probability that gives maximum expected reward in a given state or (2) select a random action with ϵ probability [149]. The parameter ϵ enables us to adjust the balance between exploitation (i.e. keep using a known action that gives local maximum reward) and exploration (i.e. go beyond the current optimal action and investigate broader solution space). The value of ϵ can be specified after we gain a better understanding of task states and possible actions from answering RQ1-RQ3.

The Algorithm1 presents the Q-learning process for a given task in one round of iteration. It usually takes a relatively large amount of iterations (e.g., over 100 iterations) for a Q-learning algorithm to converge to good performance (in this case, a smaller number of steps for completing a task) [97]. In each round of *iteration*, we run

10 *episodes*. This is because we have 10 unique search sessions under each task-topic combination. In each episode, we use a starting state of an authentic search session as the initial state of the simulated search episode. We terminate an Q-learning iteration and start a new iteration process once we exhaust all 10 starting states. This simulation setup ensures that the simulated episodes and authentic search sessions have the same starting point and thereby enables us to compare the performance of simulated search episodes with that of participants' search sessions. To further clarify the details of the proposed Q-learning-based simulation mechanism, here we present a hypothetical example about *Alice's search experience*:

Alice needs to explain the relationship between several key facts about coelacanths and conservation efforts in her article. To do this, she needs to search for useful information on the Web. Here we assume that Alice has an intelligent search system which can perfectly infer her current task state based on a variety of available signals. Alice starts with an *exploration state* as she wants to explore the general topic first before deciding the specific problems she wants to dig into. However, with limited topical and domain knowledge, the only possible query that Alice can think of at this point is "*the coelacanth*", which is not quite effective because the results generated from this query would be too broad and most of the retrieved Web pages would be useless for the task. In this context, Alice experiences the classical problem of anomalous state of knowledge (ASK) [9]: when she tries to look for information that is useful for the task at hand, she does not know what exactly she is looking for. Under this circumstance, according to the dynamic Q function, the system picks an action a_1 (e.g., an unique combination of parameter values for a predefined search path ranking algorithm) given the exploration state, which ranks the query "*coelacanths fish and fossil*" and the associated SERP and content pages as a portfolio on the top. This slice of search path sp_1 was generated by another user, John, under the same type of task state (i.e. exploration state). The associated SERP contains ten results on the first page. Two of the five clicked results are actually useful for the task. At the same time, Alice enters into the next state, *exploitation state*, which means that Alice wants to dig deeper into a specific problem or question within the broad topic. In simulation (see Figure 3.6), we simulate the next task state based on state transition probabilities we learned from answering the RQ2. The action a_1 and reward r_1 are recorded by the search system for updating Q function and policy of recommendation in the backend. The updated Q function will be used to guide the next round of action selection.

Now Alice is in her second state, exploitation state. Given this new state, the search system goes back to the (updated) Q function and picks a new action a_2 according to the value of state-action pairs. The action a_2 leads to a new slice of search path sp_2 from another user, Cathy, under the same type of task state. The recommended search path sp_2 is a new portfolio that consists of a more specific query "*why do coelacanths live 200-400 meters below the surface*" and the associated Web pages. After the action is taken and the associated reward is generated, the search system updates Q function again according to the action a_2 and the associated new reward r_2, and Alice moves on to the next task state (Again, selected or simulated based on state transition probabilities extracted from the collection of all authentic search sessions).

After several rounds of interaction with the search system, Alice successfully gathers all useful information she needs for completing the task and terminates the search session. Suppose the total reward from the obtained useful information is R. How do we evaluate the effectiveness of state-based system support?

To do this, we assume that there is a parallel universe where Alice does not have the state-aware intelligent search system. After she completes her search session for the same complex search task, the total reward she obtains from the retrieved information is R^*. Here we use R^* (which is generated from Alice's original search episode) as our baseline for system evaluation. The improvement achieved by the state-based search support can be measured by $\Delta R = R - R^*$.

Note that although the example explained above is a hypothetical one, the search task regarding coelacanths is a real task that we designed and applied in the Lab study 1: information seeking intention (ISI) study. Also, all of the queries mentioned in the example are extracted from real search logs generated by our study participants.

Note that in this dissertation study, we only use the dataset from ISI study in Q-learning for three main reasons: 1) ISI study has specified task completion requirements for all four tasks, which allows us to properly define rewards and determine the termination point for each iteration of learning and simulation. However, in the P-H study, search tasks do not have predefined task completion requirements and there are no clear indicator for measuring rewards; 2) ISI study involves four task types and two different topics, which enables us to test the state-based Q-learning approach in a variety of contexts; 3) A relatively small amount of states (four intention-based task states from the answer to RQ1) allows us to better examine the Q values of state-action pairs and update the entire policy of Q-learning algorithm in a more efficient manner.

3.5.4 Model evaluation measures

To evaluate the simulated search sessions generated by our Q-learning algorithm in a comprehensive manner, we employ multiple evaluation measures which address different aspects of task-based search interactions.

- **Number of query segments**. Number of steps or query segments is the main measure for our evaluation. Using rewards annotated based on users' bookmarked pages to iteratively update the Q-learning algorithm, we aim to reduce the steps (i.e. query segments) needed for completing the search task at hand and improve the efficiency of search. To evaluate our model at different iteration points, we compare the length of simulated search paths with the average length of users' search sessions.

Although our Q-learning algorithm is trained to exploit useful resources and improve the efficiency (reducing the steps needed) of completing complex tasks, we also care about the exploration aspect of search along the way of simulation. Thus, we add three additional measures to enrich our evaluation framework of simulated search episodes:

- **Number of relevant pages visited per query segment**. Since we do not have TREC-style relevance judgment in our user study dataset, we define page relevance based on an assumption used in previous studies (e.g., [54, 55]): If a content page is clicked by at least two different users in different sessions, then the content page is relevant to the topic.

- **Number of unique content pages visited per query segment**.

- **Average likelihood of discovery (LD) of each clicked page**.

The likelihood of discovery (LD) score of a visited Web search result is defined by Shah and Gonzalez-Ibnez [122] as follows:

$$LD(p) = \frac{|C(p)|}{|C|} \tag{3.3}$$

Where $C = \{c | c \in all\_Web\_page\_clicks\}$ and $C(p) = \{Clicks\_on\_thepage\_p\}$. The LD score is useful in measuring the contribution of search recommendation in supporting exploration in search sessions and the construction of new knowledge [122].

3.6 Back to the fundamentals: Assumptions in this study

Before moving on to the actual data analysis and result presentation, it is critical to first carefully examine the ground upon which we build the entire dissertation study. However, it is impossible for us to exhaust all possible assumptions and to discuss every one of them. Hence, in this section, we seek to identify the most important assumptions behind our methodology and explain (1) why we need them and (2) what they mean for the study (e.g., potential benefits, limitations, and constraints). It is necessary to keep these assumptions in mind when interpreting the results and findings.

- **Assumption 1**: *Users do not change their local information seeking intention(s) within a query segment. Changes of information seeking intention(s) only happens when the associated query changes.* This assumption allows us to use query segment (which starts from a query and includes the browsing and search result examination behaviors associated with the query) as the basic unit for intention and task state analysis. However, it is worth noting that in a long query segment with rich browsing activities, a user might change his or her local intentions during the query segment, without reformulating a new query. The sequence of intention transitions within (long) query segments are not studied in this dissertation work. Although our method cannot explicitly capture the changes of intentions within query segments, it still allows participants to identify multiple intentions when necessary or applicable. The sequence of these intentions within individual query segments is not the focus of this dissertation study or part of the research questions we seek to address here. For extracting task state categories and simplifying the process of analysis, we decided to use unique combinations of intentions (as different dimensions of users' intention space) to represent different intention states.

- **Assumption 2**: *In this study, the observed difference in task states and state transitions is associated with task type, not the difference in users' knowledge backgrounds.* Users' topic and domain knowledge can significantly affect their task perceptions as well as the process of doing complex tasks [85, 160]. To address this issue, in both of the user studies, we selected topics that our participants are unlikely to be familiar with, aiming to control the potential effects of the variations in participants' topic knowledge and familiarity. However, it is likely that the generalizability and applicability of our state-based recommendation approach is constrained by the topic and the population where we recruited our participants. Hence, a large scale experiment based on a variety of topics and communities will be a critical next step for this line of research.

In addition to the above assumptions made for data collection in lab studies, we also made additional assumptions for running the Q-learning-based simulation of search recommendations.

- **Assumption 3**: *The totality of all participants' search paths presents the accessible solution space or space of methods behind a given complex search task.* Build upon this assumption, we build reinforcement learning algorithm in order to find the optimal search path that generates maximum amount of expected rewards. Also, this assumption allows us to use existing collection of search paths to simulate the responses (rewards and next task state) from the dynamic search environment. It is possible that for the given search task, there is a better, unexplored search path which is not included in our collection. Nevertheless, we can still learn potentially useful search paths and tactics from existing users' search explorations, which is also meaningful for improving adaptive search supports and recommendations.

- **Assumption 4**: *Users always accept and use the state-based adaptive supports provided by interactive search systems.* This assumption allows us to fully evaluate the value of state-based search supports and simplify the evaluation process. For future HCI studies on the issue of state-based search recommendations, we need to go beyond this assumption and investigate the contextual factors that may significantly affect users' acceptance of system recommendations and interventions. Liu et al. [91] found that the behavioral impacts of search interventions and persuasions are not only affected by the contents, but also by the sources of recommendations (i.e. who is recommending these contents). This problem of system and recommendation acceptance is not the focus of this dissertation work.

3.7 Summary

Figure 3.7 summarizes the research questions, methods, and types of data used in this dissertation study. In summary, this dissertation has two main goals: (1) *understanding states of complex search tasks*; (2) *leveraging the knowledge of task state in supporting users*. To reach these two goals, we propose four research questions. RQ1 and RQ2 focuses on state identification and modeling of state transitions. To address these two questions, we utilize users' original annotation on information seeking intentions (in Lab study 1), search problems and preferred system supports (in Lab study 2) and seek to extract meaningful task state categories from the annotation dataset through K-modes clustering. Then, we use external judgment method to test the validity of the task state types generated by the clustering algorithm. RQ3 takes a step forward and speaks to the connection between observable search behaviors and implicit task states. We first use correlation analysis methods to statistically test the behavior-state connection, and then investigate the extent to which we can infer and predict task state from search behavior using supervised learning (classification) models.

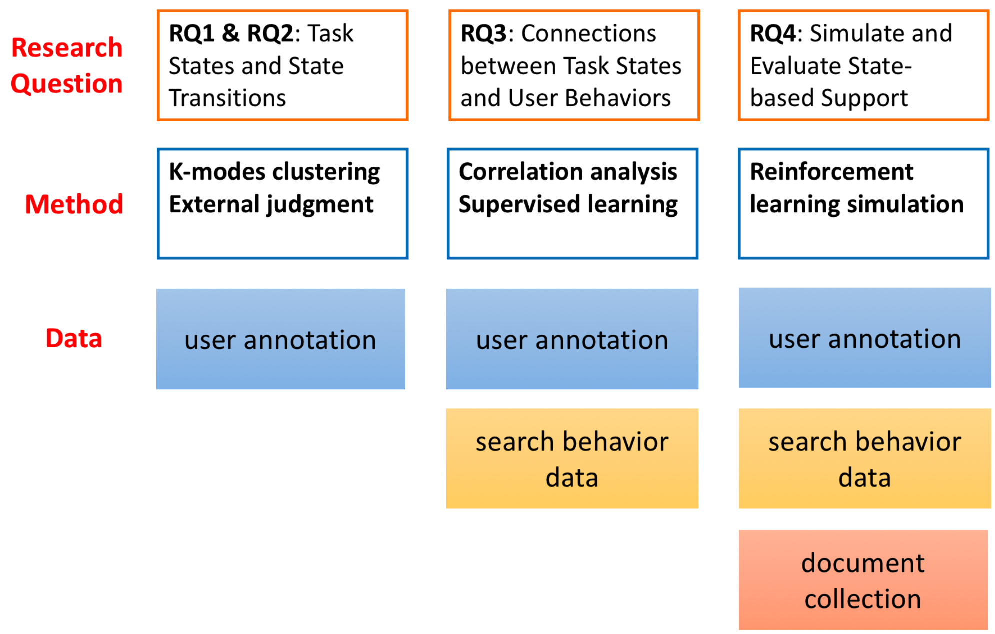


Figure 3.7: Summary of research questions, methods, and types of data used.

RQ4 speaks to the second main goal of this dissertation study. Specifically, to answer this research question, we utilize all types of data we collected from user studies (i.e. user annotation, search behavior data, collection of documents retrieved) and run simulations using reinforcement learning approach. Specifically, we seek to simulate the response from the search environment (i.e. reward, next task state) and evaluate the simulated state-based search path recommendations by comparing them with users' original search sessions.

This chapter introduced the ways in which we define task states and utilize the knowledge of states in dynamically supporting users engaging in complex search tasks. The user study procedures, tools employed, and the collected datasets are explained in detail. Through answering the four research questions, this dissertation aims to (1) enhance the understanding of complex search tasks through studying the transitions of task states and examining the connections between the states and users' search behaviors; (2) leverage the knowledge of task states in simulating and evaluating adaptive system recommendations in complex search tasks of different types.

Chapter 4

Results

To answer the proposed three RQs, we analyzed the search behavioral data and user annotations collected from ISI and P-H studies. In ISI study we collected data from 693 query segments generated by 40 participants in 80 task-based search sessions. The P-H 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 P-H 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.

4.1 RQ1: Identify task states

The identification of task states started with K-modes clustering analysis. Before that, we determined the optimal number of clusters based on the results from elbow analyses and cluster distributions. We extracted four clusters as separate task states from the ISI dataset and six clusters from the P-H dataset. The clustering analysis for P-H 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.

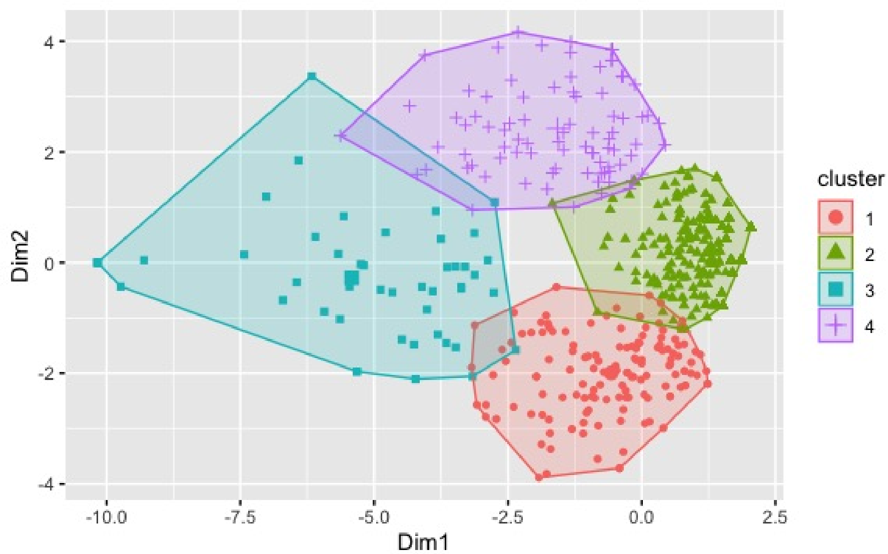


Figure 4.1: Clusters extracted from intention annotations.

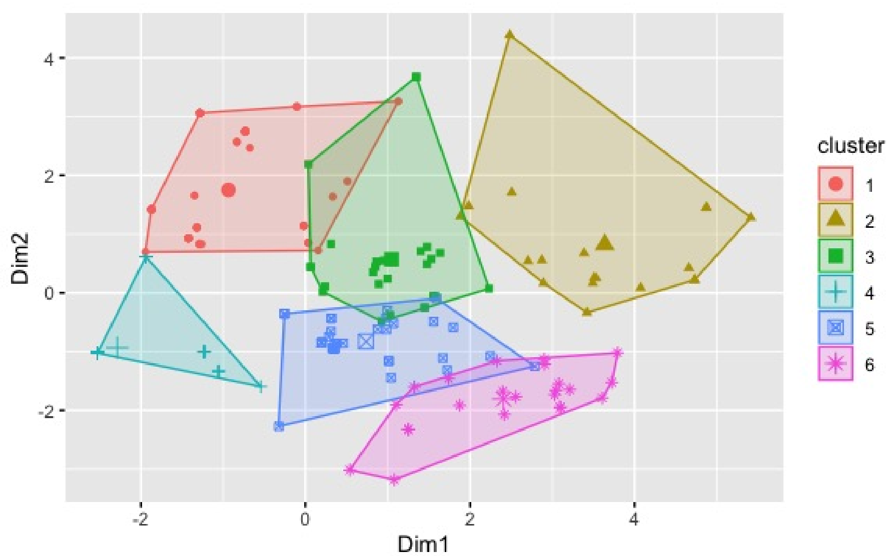


Figure 4.2: Clusters extracted from problem-help annotations.

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** (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 certain topic in mind and they try to follow the current search path, keep exploiting the information patch at hand and search for more relevant documents, pages or other information items.

- **Known-Item** (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 target item(s) in mind to guide their search interactions. These specific items may come from previous interactions and are not necessarily in users' minds before search.

- **Exploratory** (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** (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.

Note that under each task state, we list the most frequent information seeking intentions with the associated percentages. For instance, in the *Exploratory* state, the percentage of the intention of *identifying something to start searching* is 100%, indicating that this intention occurred in all query segments under this state or cluster. 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 occurring 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 [9] (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%).

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 (P-H-based state); 2) annotator A and clustering algorithm: 0.716 (ISI-based state), 0.717 (P-H-based state); 3) annotator B and clustering algorithm: 0.744 (ISI-based state), 0.682 (P-H-based state). The Cohen's Kappa agreements in all pairs are above 0.65, which is considered *substantial* agreement [76]. 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) [76], indicating that inferring implicit task states from search interactions is not an easy job (even for human annotators).

Table 4.1: Differences in intention frequency distributions across task states.

Intention	Chi-squared Test	Effect Size
access website	χ^2**=42.03, p<.001**	0.246
access common	χ^2**=128.72, p<.001**	0.431
access item	χ^2**=135.91, p<.001**	0.443
evaluate best	χ^2**=116.24, p<.001**	0.410
evaluate correct	χ^2**=89.32, p<.001**	0.359
evaluate duplicate	χ^2**=15.342, p=.0015**	0.149
evaluate specific	χ^2**=180.94, p<.001**	0.511
evaluate usefulness	χ^2**=138, p<.001**	0.446
find common	χ^2**=86.39, p<.001**	0.353
find known	χ^2**=33.35, p<.001**	0.219
find specific	χ^2**=214.11, p<.001**	0.555
find without	χ^2**=12.24, p=.0066**	0.133
identify more	χ^2**=76.23, p<.001**	0.332
identify start	χ^2**=483.16, p<.001**	0.835
keep link	χ^2**=206.37, p<.001**	0.546
learn database	χ^2=7.21, p=.0.066	0.102
learn domain	χ^2**=149.4, p<.001**	0.464
obtain part	χ^2**=56.79, p<.001**	0.286
obtain specific	χ^2**=382.4, p<.001**	0.743
obtain whole	χ^2**=35.20, p<.001**	0.225

Note: The statistically significant results are boldfaced. The threshold of effect size (Cramer's V) are: 0.1 (small effect size), 0.3 (medium effect size), and 0.5 (large effect size) [32].

The frequency distributions of 20 information seeking intentions under different task states are presented in the section 7.2 of the Appendices chapter. To shed light on the difference across different states in terms of intention distribution, we conducted chi-square test to examine the association between the proportion of information seeking intention states (i.e. present versus absent) and intention-based task states (see Table 4.1). We sought to capture the statistically significant differences in intention frequency distributions across task states. Similarly, we also tested the association between P-H state and the frequency distribution of each problem or help item (see Table 4.2). The results demonstrate that there are significant differences across intention-based and problem-help-based task states in almost all dimensions, and that the strengths (measured by effect size) of most state-intention and state-problem/help associations fall into the range of medium or large effects.

To further explore the boundaries and distinctions 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

Table 4.2: Differences in problem-help frequency distributions across task states.

Problem/Help	Chi-squared Test	Effect Size
P: difficult articulate	**χ^2=86.17, p<.001**	0.632
P: irrelevant result	**χ^2=59.72, p<.001**	0.526
P: lack knowledge	**χ^2=89.59, p<.001**	0.644
P: lack patience	–	–
P: credibility uncertain	**χ^2=12.55, p=.006**	0.241
P: source unaware	**χ^2=89.11, p<.001**	0.642
P: info overload	**χ^2=19.38, p=.0016**	0.300
P: source unavailable	**χ^2=48.63, p<.001**	0.474
no problem	**χ^2=107.17, p<.001**	0.704
H: page recommend	**χ^2=124.95, p<.001**	0.761
H: people recommend	**χ^2=83.67, p<.001**	0.622
H: query recommend	**χ^2=122.33, p<.001**	0.753
H: strategy recommend	**χ^2=133.73, p<.001**	0.787
H: system unsatisfactory	**χ^2=19.08, p=.0018**	0.297
no help needed	**χ^2=175.51, p<.001**	0.901

Note: The statistically significant results are boldfaced. The threshold of effect size (Cramer's V) are: 0.1 (small effect size), 0.3 (medium effect size), and 0.5 (large effect size) [32]. The problem of "lacking patience" was not selected by any participant in the lab study.

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 P-H studies, we employed *Dunn's test* with Benjamini-Hochberg (B-H) correction for post hoc pairwise analyses.

Table 4.3: Behavioral variations across different task states: Median (IQR) (*: $p < .05, ** : p < .01$).

behavior	Exploit	Known	EXplore	Learn	Dunn's posthoc test
querylength*	4(3)	4(3)	3(4)	3(3)	$E > EX^*, K > EX^*, E > L^*, K > L^*$
dwell_SERP**	7.2(10)	7.3(10)	6.2(7)	4.9(3)	$K > EX^*, E > L^*, K > L^{**}$
avg_dwell_content**	7.9(14)	13.5(15)	8.9(17)	13(15.9)	$K > E^*, K > EX^*, L > E^*, L > EX^*$
N.content**	5(2)	4(3)	4(4)	3(2)	$E > L^*, K > L^*$
total_dwell_content**	33(69)	67(92)	35(28)	54(83)	$K > E^*, K > EX^*, L > E^{**}, L > EX^*$
N.clicks**	2(3)	3(4)	3(4)	4.5(5.5)	$L > E^{**}, L > EX^*, L > K^*, EX > E^*$
N.bookmark**	0(1)	0(1)	0(1)	1(2)	$L > E^*, L > K^*, L > EX^*$

Table 4.3 presents the results of Kruskal-Wallis and Dunn's post hoc tests (with B-H correction) on the behavioral variation across different intention-based task states. Since the data was not normally distributed, table 4.3 reports the medians and inter-quartile ranges (IQR) (instead of means and standard deviations) of behavioral measures. 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). These results demonstrate that the intention-based task states are closely associated with participants' selections of search tactics in local search steps.

Table 4.4: Behavioral variations across different intention-based task states: Median (IQR) (*: $p < .05, ** : p < .01$).

behavior	IO-P	A-LT-PE	A-SU-M	NP	LT-M	SU-QU
querylength*	3.5(2)	4(1)	4(2)	4(3)	4(1)	4(2)
dwellSERP**	34.5(26)	81(68)	89(84)	37(38)	41(39)	42(43)
avg_dwell_content	17(44)	21.5(23.67)	35(44)	11.63(44)	36(48)	13.25(23.91)
N.content*	1(2)	2(1)	0(1)	1(2)	1(1)	1(1)
total_dwell_content**	18.5(58)	43(68)	67(95)	21(45)	40(48)	14.5(53)
N.clicks**	1.5(3)	2(2)	1(2)	2(3)	1(2)	1(2)
N.bookmark**	0(2)	0(0)	0(1)	1(1)	0(1)	0(1)

Table 4.5: Behavioral variations across different problem-help-based task states: Dunn's posthoc test with B-H correction ($* : p < .05, ** : p < .01$).

behavior	Dunn's posthoc test
querylength*	$ASK - SU - M > IO - P^{**}, ASK - LT - PE^{*}, NP^{*}, SU - QU^{*}$
dwell_SERP**	$ASK - LT - PE > IO - P^{**}, NP^{**}, LT - M^{*}, SU - OU^{*}; ASK - SU - M > NP^{*}, SU - OU^{**}$
avg_dwell_content	N.A.
N.content*	$ASK - LT - PE > IO - P^{*}, ASK - SU - M^{*}, NP^{*}, SU - QU^{*}; ASK - SU - M > NP^{*}$
total_dwell_content**	$ASK - LT - PE > IO - P^{*}, ASK - SU - M^{*}, NP^{*}, LT - M^{*}, SU - QU^{*}$
N.clicks**	$IO - P > ASK - SU - M^{*}; ASK - LT - PE > ASK - SU - M^{*}, LT - M^{*}, SU - QU^{*}$
N.bookmark**	$NP > ASK - LT - PE^{**}, ASK - SU - M^{*}, LT - M^{*}, SU - QU^{*}$

Tables 4.4 and 4.5 illustrate the behavioral variations across different problem-help states. The results indicate that 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 analyses help clarify the boundaries between different task states at both cognitive and behavioral levels and thereby pave the way for predicting task states from search behavioral signals (RQ3).

4.2 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 4.3, 4.4, 4.5 and 4.6 illustrate the state transition patterns (including the frequency distributions of start states and end states) of four types of complex search tasks assigned in the ISI study, and Figures 4.7 and 4.8 presents the state transition patterns of the two task types in P-H study.

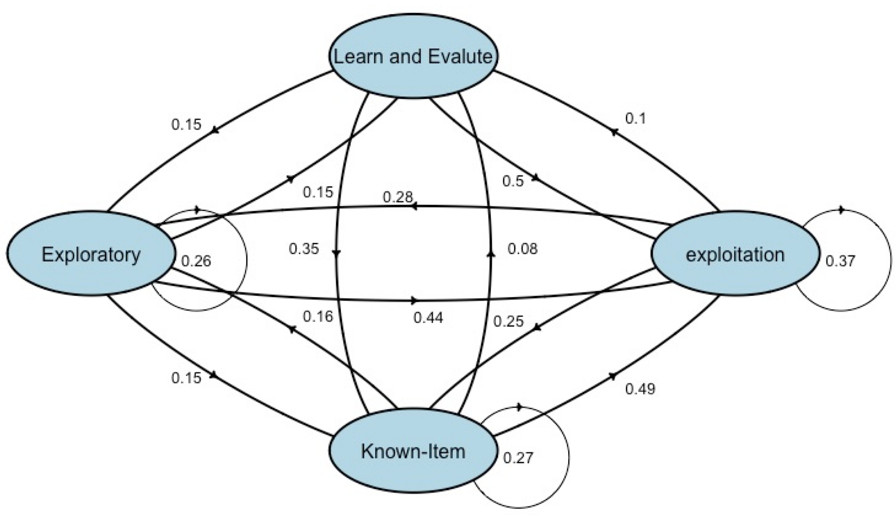


Figure 4.3: ISI copy editing task (factual specific).
Start states distribution: *Exploitation*: 11.1%; *Known-item*: 16.7%; *Exploratory*: 41%; *Learn and Evaluate*: 22.2%; **End state distribution**: *Exploitation*: 54.6%; *Known-item*: 18.2%; *Exploratory*: 9%; *Learn and Evaluate*: 18.2%.

Overall, our results demonstrate that the process of doing a complex search task

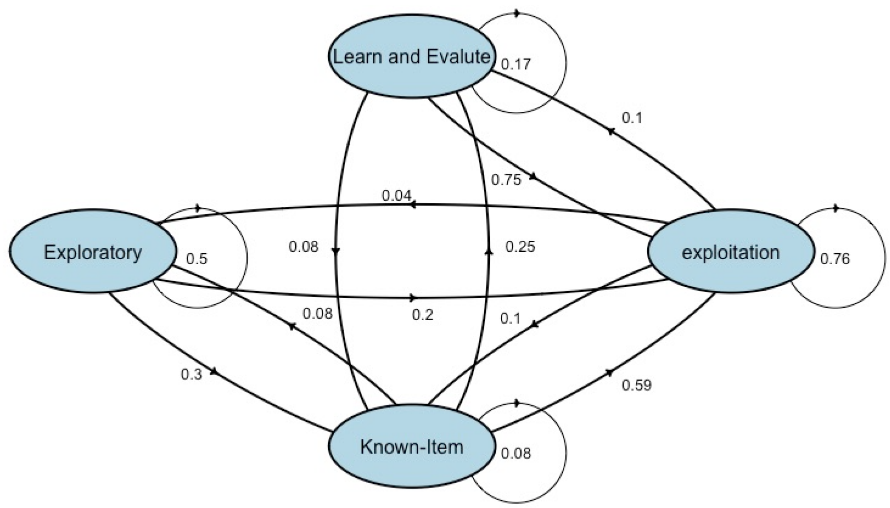


Figure 4.4: ISI story pitch task (factual amorphous).
Start states distribution: *Exploitation*: 13.6%; *Known-item*: 22.7%; *Exploratory*: 50%; *Learn and Evaluate*: 22.7%; **End state distribution**: *Exploitation*: 61.1%; *Known-item*: 22.2%; *Exploratory*: 0%; *Learn and Evaluate*: 16.7%.

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. [78]) 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 an ill-defined goal.

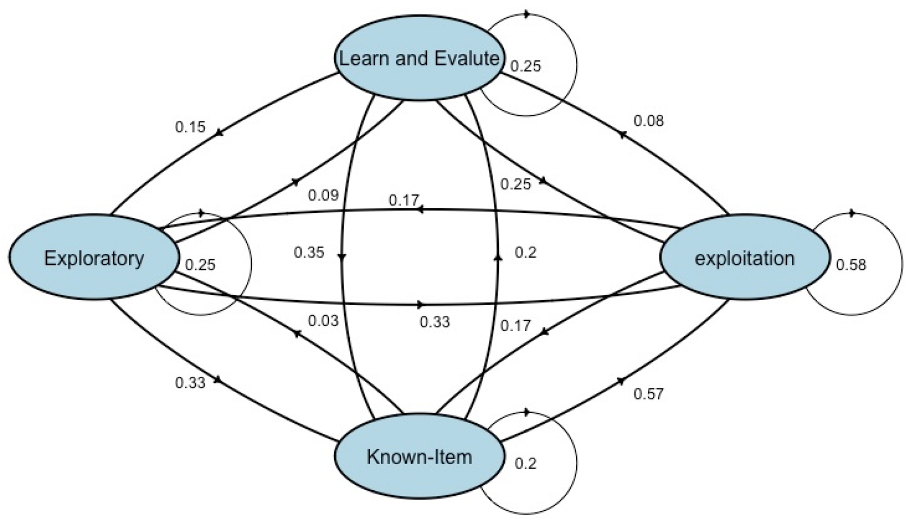


Figure 4.5: ISI relationship task (intellectual amorphous).
Start states distribution: *Exploitation*: 5.6%; *Known-item*: 22.2%; *Exploratory*: 50%; *Learn and Evaluate*: 22.2%; **End state distribution**: *Exploitation*: 61%; *Known-item*: 16.7%; *Exploratory*: 16.7%; *Learn and Evaluate*: 5.6%.

The problem-help-based task states offers us a different perspective on the dynamic nature of complex search tasks. Figures 4.7 and 4.8 illustrate the problem-help state transition patterns. Since the problem-help approach produced six separate task states, to improve the clarity and readability of these two diagrams, 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 combinations of query terms is a major intellectual 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

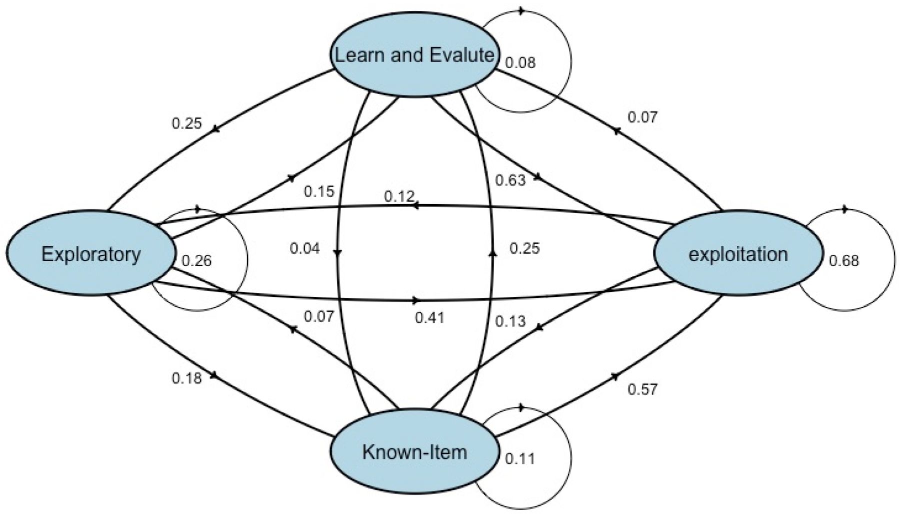


Figure 4.6: ISI interview prep task (intellectual amorphous).
Start states distribution: *Exploitation*: 0%; *Known-item*: 9.1%; *Exploratory*: 68.2%; *Learn and Evaluate*: 22.7%; **End state distribution**: *Exploitation*: 81.8%; *Known-item*: 9.1%; *Exploratory*: 0%; *Learn and Evaluate*: 9.1%.

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.

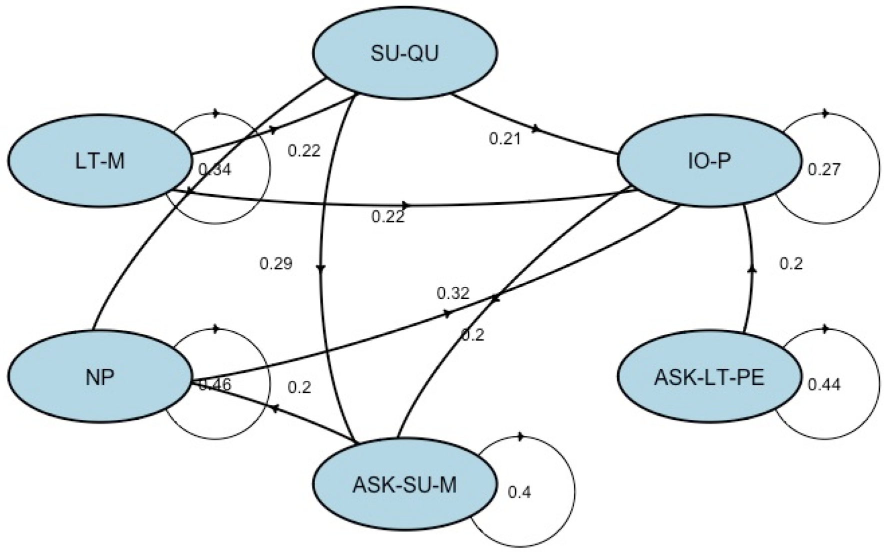


Figure 4.7: P-H cognitive task (intellectual amorphous).

Start states distribution: *IO-P*: 11.5%; *ASK-LT-PE*: 15.4%; *ASK-SU-M*: 7.7%; *NP*: 30.8%; *LT-M*: 19.2%; *SU-QU*: 15.4%; **End states distribution**: *IO-P*: 23.1%; *ASK-LT-PE*: 19.2%; *ASK-SU-M*: 15.4%; *NP*: 19.2%; *LT-M*: 15.4%; *SU-QU*: 7.7%.

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 the probability of transition from ASK-related state to NP in 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.

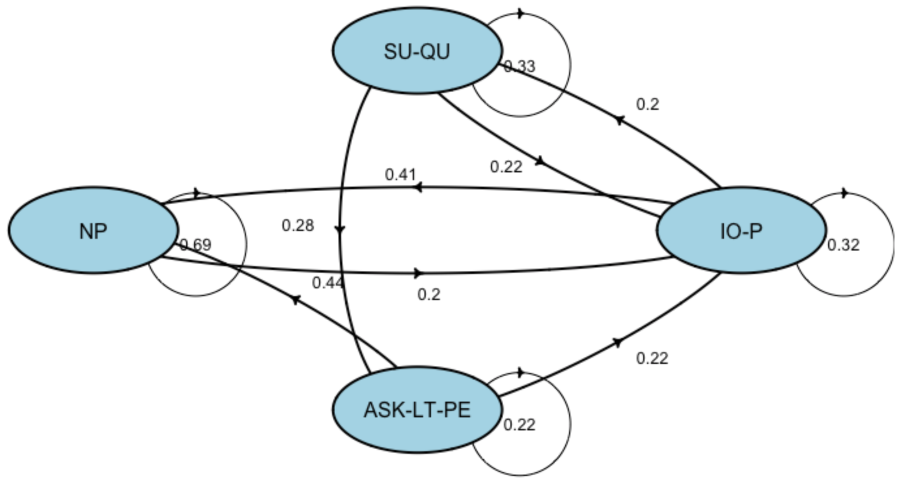


Figure 4.8: P-H social task (intellectual amorphous).

Start states distribution: *IO-P*: 21.7%; *ASK-LT-PE*: 8.6%; *NP*: 52.3%; *SU-QU*: 17.4%; **End states distribution**: *IO-P*: 25%; *ASK-LT-PE*: 16.7%; *NP*: 50%; *SU-QU*: 8.3%.

In addition, the dynamic nature of complex search tasks can also be partially revealed through analyzing how the relative frequencies of different task states vary across different stages of task processes. Comparing the temporal variations of the relative frequencies of intention-based and problem-help-based task states can enhance our understanding about how the difference between search tasks of different types is manifested at cognitive level.

The frequency densities of intention states are presented in Figures 4.9, 4.10, 4.11 and 4.12. In these figures, each dot presents a query segments occurring at a particular point of the associated search session. Overall, participants tended to use more queries in the factual specific task (i.e. copy editing) compared to the three goal-amorphous tasks. In the two cognitively challenging, intellectual-amorphous tasks, the density plot peak points of known-item state and learn-and-evaluate state show up at early stages of search sessions (query percent < 0.3). This result indicates that in these two tasks, participants frequently searched for easily accessible, known items and sought to learn and evaluate information at the beginning of search sessions. In contrast, in the two factual tasks, the curves of frequency density distribution are relatively flat and the density plot peak points appear a bit later compared to the two intellectual tasks. In addition, in the two intellectual amorphous tasks, due to the lack of clear informational cues and search paths, participants tended to do exploratory search at early search stages, with the peak point of exploratory-state frequency density plot occurring around the 0.25 query percent point.

With respect to the temporal distribution of P-H states, the peak point of SU-QU state (frequent problem: unaware of useful information sources; help needed: query recommendation) shows up early in search episodes (around 0.25 query percent) in the cognitive task. In the social task, the SU-QU state points are evenly distributed across the entire search process. In addition, due to task complexity and lack of topical knowledge, participants encountered ASK-related P-H states (frequent problem: do not know how to express information needs with search queries) more often in the cognitive task. Since the topic and work task context of the social search task were more familiar to our participants (i.e. college students at Rutgers University), they encountered less obstacles in this task. As a result, we observed a higher frequency of NP cases and the absence of ASK-SU-M state (frequent problem: ASK issue and the problem of not knowing useful information sources). Also, there was a low frequency of occurrences of the ASK-LT-PE state at the first half of search sessions in the social task. This may be because participants tended to utilize all known informational cues in formulating queries and to collect all "low-hanging fruits" at the early stages of search. After they exhausted all easily accessible information, the frequencies of ASK and information-overload problems gradually increased as the search process proceeded. We found similar patterns of temporal changes in the cognitive tasks, but the overall frequencies of these problem states (information overload, ASK, and lack of topical knowledge) were higher than that of the social task.

In summary, the differences in state transition probabilities and temporal distributions of states shed light on the divergences between complex tasks of different types from a process-oriented perspectives. The knowledge of these differences can complement our understanding of the nature of complex search tasks and push us closer to predicting task states and developing state-aware adaptive search recommendations and interventions.

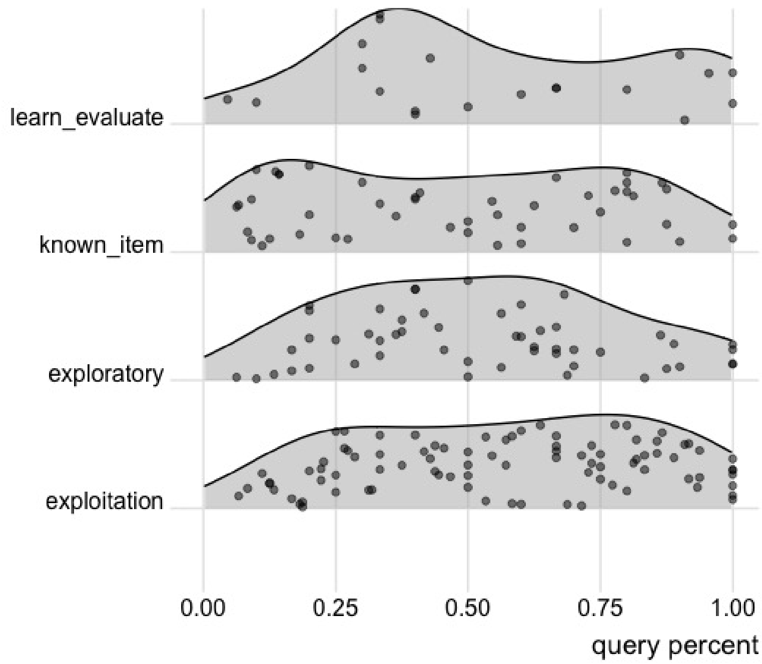


Figure 4.9: Frequency density of intention states at different search stages: CPE. Each dot represents a query segment.

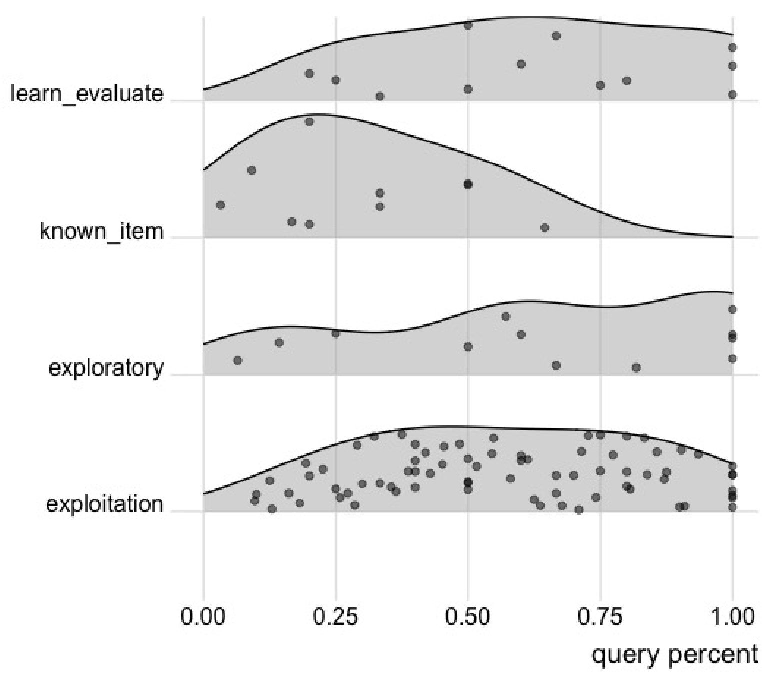


Figure 4.10: Frequency density of intention states at different search stages: STP.

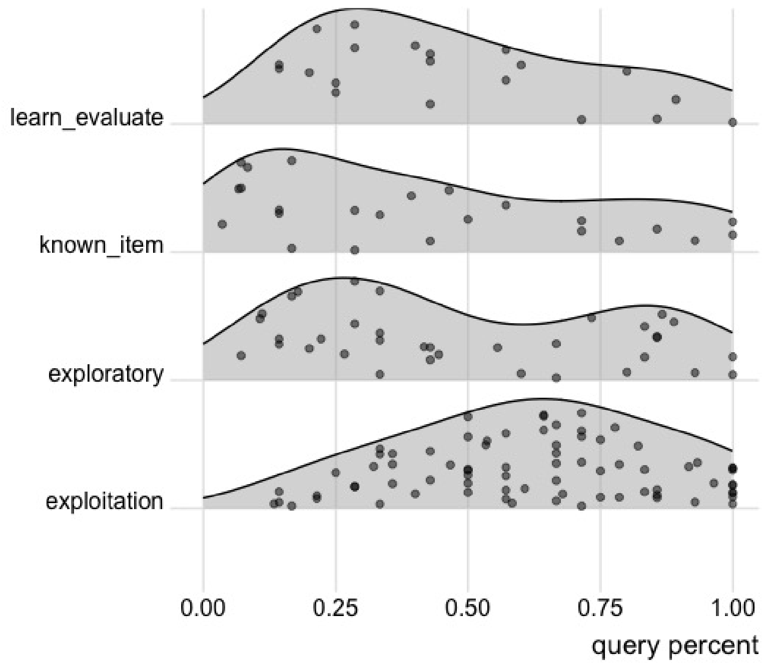


Figure 4.11: Frequency density of intention states at different search stages: REL.

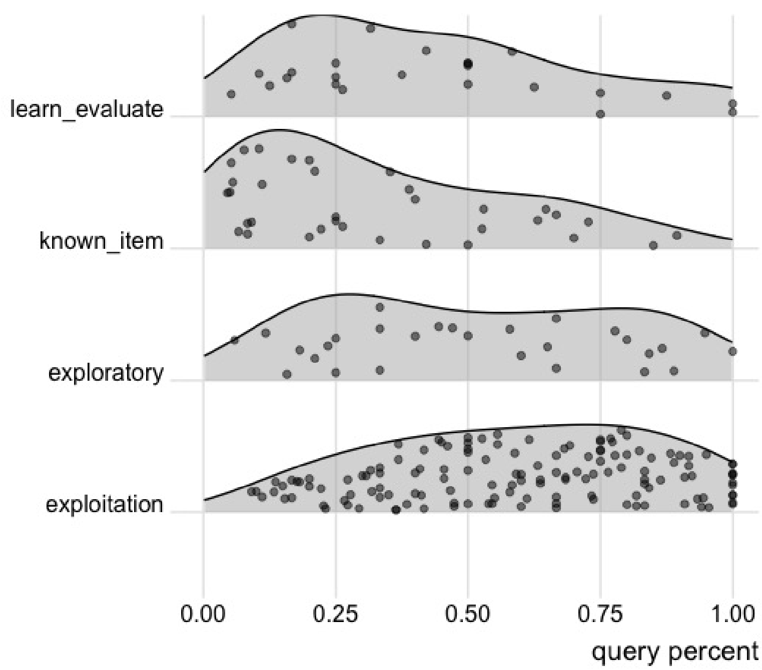


Figure 4.12: Frequency density of intention states at different search stages: INT.

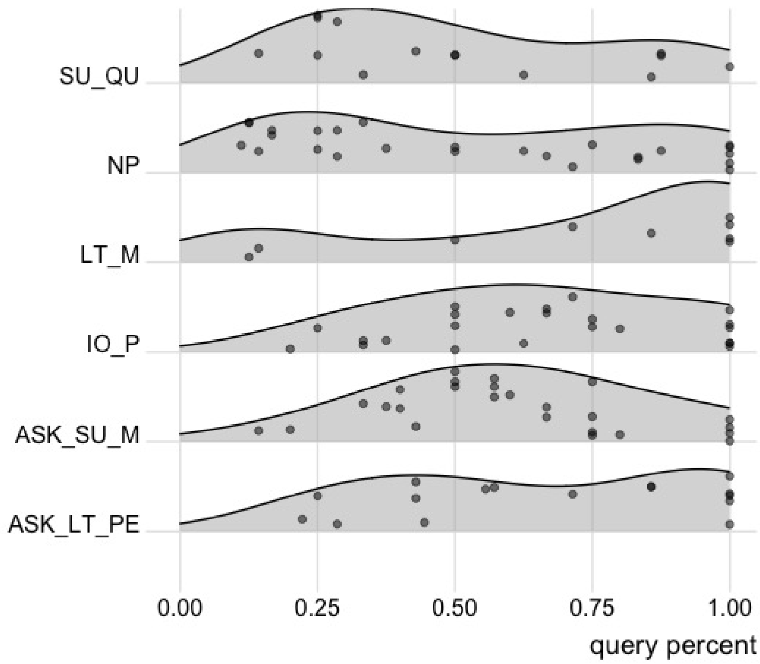


Figure 4.13: Frequency density of P-H states at different search stages: Cognitive.

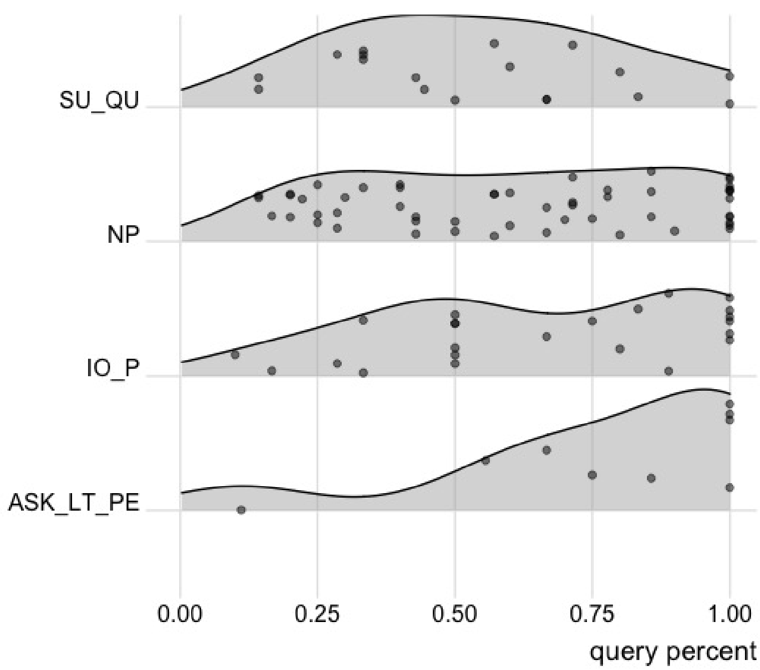


Figure 4.14: Frequency density of P-H states at different search stages: Social.

4.3 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 implicit task states from observable behavioral signals. To answer RQ3, we built several classifiers based upon the behavioral features introduced in the section 3.4. 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. Also, it would lead to difficulty for developing useful models for predicting these states when the available search behavioral data is limited. We therefore did two parts of analysis: 1) developing classifiers for six-state prediction, and 2) collapsing six states into larger, meaningful state categories or types and building models for P-H state type prediction. The method of collapsing detailed state categories into broader types or groups is applied in a variety of IR prediction analyses (e.g., [142, 82, 160]) and allows us to examine the predictability of problem-help states at different levels of granularity.

Based on the distribution of specific problem and help features, we collapsed six specific P-H states into the following three types: 1) P-H state with *one dominant type of preferred support*, including IO-P (page recommendation), ASK-LT-PE (people recommendation), and SU-QU (query recommendation); 2) P-H state with *multiple types of preferred support*: including ASK-LP-SU-MULTI (Web page recommendation: 88%; query recommendation: 96%; search strategy recommendation: 92%) and LT-MULTI (Web page recommendation: 89%; people or expert recommendation: 89%; query recommendation: 78%; search strategy recommendation: 100%); 3) *No-problem-encountered/No-support-needed* state: NP. In this state, users did not encounter any major problem that hinders the ongoing Web search and task completion process.

We trained and evaluated machine learning classifiers with an 80/20 split on training/testing data and compared them with two baselines: 1) random baseline and 2) most frequent labeling baseline. The accuracy scores for prediction analyses are presented in Table 4.6 (ISI state prediction), Table 4.7 (P-H state prediction).

Table 4.6: 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.

The findings presented in Table 4.6 and Table 4.7 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 using behavioral data from current query segment alone 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. In P-H state type prediction (see Table 4.8), we found similar patterns of prediction results: 1) current-segment-based classifiers usually achieve best performance; 2) models built on previous-session features alone also reach better accuracy level than the best baseline. In general, our models achieved better performances in this "lower resolution" prediction, with the best performer reaching almost 70% accuracy in state type prediction. The evaluation results associated with other measures (e.g., AUC, F1 score) are provided in the appendix. 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 task state transitions in search interactions and to develop state-aware adaptive system supports.

Table 4.7: 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

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

In the next section, based on the knowledge we learned about task states, state

Table 4.8: Accuracy score of P-H state type prediction.

Classifier	Current seg.	Prev. session	All data
Logistic Regression	0.518	0.431	0.486
Support Vector Machine	0.670**	0.594*	0.693**
XGBoost	0.692**	0.621**	0.680**
Random Forest	0.684**	0.629**	0.681**
Decision Tree	**0.696****	0.617**	0.689**
Most Frequent	0.477	0.477	0.477
Random	0.337	0.337	0.337

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

transition probabilities and users' search interactions from previous analyses, we build a model using Q-learning algorithm and seek to simulate and evaluate state-based query segment recommendations. This part of modeling allows us to answer the proposed RQ4 and also demonstrate the potential value of state-based framework in supporting users engaging in complex search tasks.

4.4 RQ4: State-based search path recommendation

This section presents to results we obtain through the Q-learning process explained in the *Methodology* chapter. As it is discussed in previous sections, to answer RQ4, we apply the state-based framework to building Q-learning algorithms and aim to utilize the knowledge we learned from answering RQs 1-3 in supporting users engaging in complex search tasks of different types. The following sections provide more details regarding the evaluation of our Q-learning algorithm in terms of different dimensions of search interactions.

4.4.1 Number of steps needed for task completion

To evaluate the state-based model, we build Q-learning algorithm using the states, actions, and rewards defined based on the nature of ISI tasks (i.e. copy editing, story pitch, relationship identification, and interview preparation) and simulate search sessions consisting of individual recommended query segments. Our main goal here is to reduce steps (query segments) needed for completing the given task (satisficing all the requirements specified in task descriptions) and thereby improving the efficiency of search interaction.

We evaluate our model by comparing the number of steps in simulated search sessions with two baselines: 1) *number of steps in original search sessions*; 2) *number of steps needed for actual completion*, which refers to the minimal number of steps needed for satisficing the specified task requirements. This measure was computed at that point in the "original search" when pages containing enough answers to the search task had been displayed to the user (i.e. the task-based reward requirement defined in the section 3.5.2 is satisfied from the "information supply" side: the values in all elements of the corresponding task-based cumulative vector are greater than zero), regardless of users' search strategies or knowledge states (e.g. whether a user fully understood the information presented on a visited page). Thus, number of steps needed for actual completion is essentially the minimal number of query segments needed for presenting "just enough" relevant information within a given search session. The minimum requirement for task completion varies across different task types and thus were operationalized using different vectors (see section 3.5.2).

Recall that in the ISI user study, participants were asked to search for information that might be useful for completing the predefined task. Participants were not required to do anything beyond the search task requirements. However, we observed varying levels of divergence between the the length of original search session and the length of actual completion sessions in all task-topic combinations: participants tended to continue their search processes and issue more queries when they already met the requirements of task completion, due to four possible reasons: 1) participants were not sure if the bookmarked pages include enough information for meeting the minimum requirement of search tasks; 2) there was still plenty of time left before the participant hit the predefined session termination point (20 minutes), which encourage him or her to revisit the clicked and bookmarked content pages and verify the information they saved; 3) in intellectual amorphous tasks, due to the difficulty of search tasks and the ambiguity of search goals, participants might need to explore more pages and learn more about the associated topic(s) in order to fully understand the information gathered in previous query segments; 4) part of the useful information on bookmarked pages (which was captured and included in manual reward annotation) was missed by participants in their original search sessions. Consequently, they might continue searching for this part of information (e.g. for confirming some of the statements in task description) that was already presented in their previously bookmarked pages. For instance, on a Google books page, a participant might be searching for information about the size of Coelacanth (which can be inferred based on the query stored in the associated URL). However, he or she might not notice that the bookmarked Google books page also contains useful information about the process of discovering Coelacanth in South Africa and Indonesia.

As it is discussed in the Methodology chapter, for every task-topic pair (e.g., CPE-Coelacanth), in each iteration, we run 10 episodes or sessions of simulation and use the starting states from all 10 authentic search sessions as the initial states of our 10 simulated search sessions. For evaluation, we compare the average number of steps in every round of iteration with the average numbers of steps from both the actual completion baseline and participants' original search sessions.

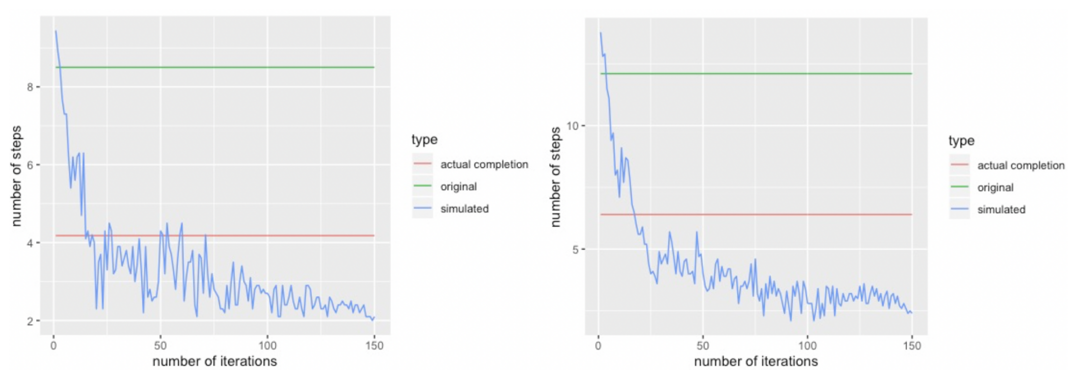


Figure 4.15: Number of steps in CPE: Coelacanth (left) and Methane Clathrates (right).

Figure 4.15 plots the performance of simulated search sessions in copy editing (CPE) tasks. We observe that the state-based simulated search path outperforms both original session and actual completion baselines before the 75th iteration in both topics. At the last two iterations, the simulated search path statistically significantly outperforms both baselines (t-test, p-value<.01) by reducing two to four steps on average for search task completion. This result indicates that through learning the connection between rewards received and the context of recommendation (i.e. state-action pairs), our state-based

Q-learning algorithm finds shorter search paths for completing CPE tasks and improves the efficiency of task-based search interaction.

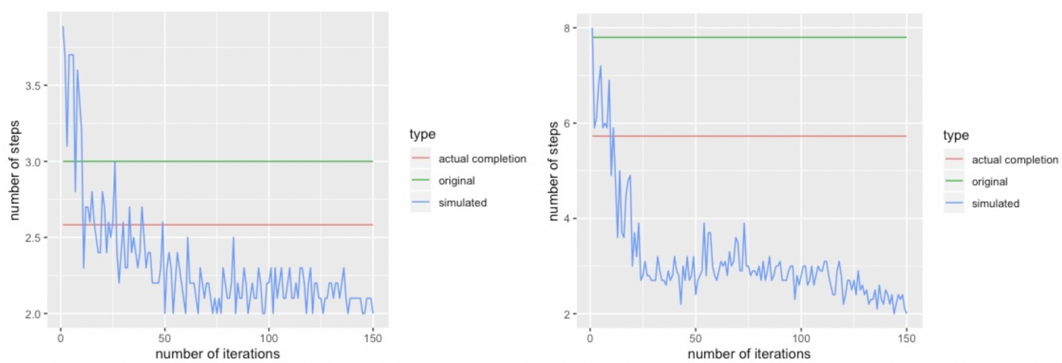


Figure 4.16: Number of steps in STP: Coelacanth (left) and Methane Clathrates (right).

With respect to the story pitch (STP)-Coelcantah task, participants' average number of steps needed for completing the task is 2.583, which does not leave much room for Q-learning improvement. Nevertheless, our state-based model still achieves slightly better performances than the actual completion baseline, with the number of steps being reduced by 0.1 to 0.48 steps on average after the 60th iteration. Compared to the STP-Coelacanth task, STP-Methane-Clathrates task appeared to be more difficult, with participants taking significantly more steps for finding and saving pages that contain six most interesting facts about the topic or research on the topic. This hypothesis is partially confirmed by our data about participants' post-search task perception: According to the ratings, STP-Coelacanth has a lower average level of perceived difficulty ($Mean\_difficulty_{Coelacanth} = 1.2, Mean\_difficulty_{Methane\_Clathrates} = 2.1; p < .05$) and a higher level of perceived task success ($Mean\_perceived\_success_{Coelacanth} = 6.5, Mean\_perceived\_success_{Methane\_Clathrates} = 4.8; p < .05$). As a result, in the Methane Clathrates task, the Q-learning algorithm achieves a much higher improvement and reduces the average length of search session by 2 to 3.6 steps after the 60th training iteration. This result demonstrates the potential value of state-based Q-learning search supports in helping users engaging in difficult, unfamiliar search tasks.

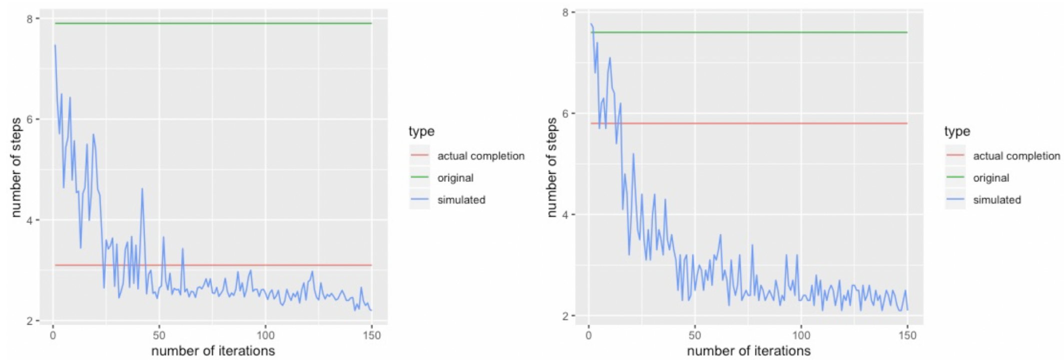


Figure 4.17: Number of steps in REL: Coelacanth (left) and Methane Clathrates (right).

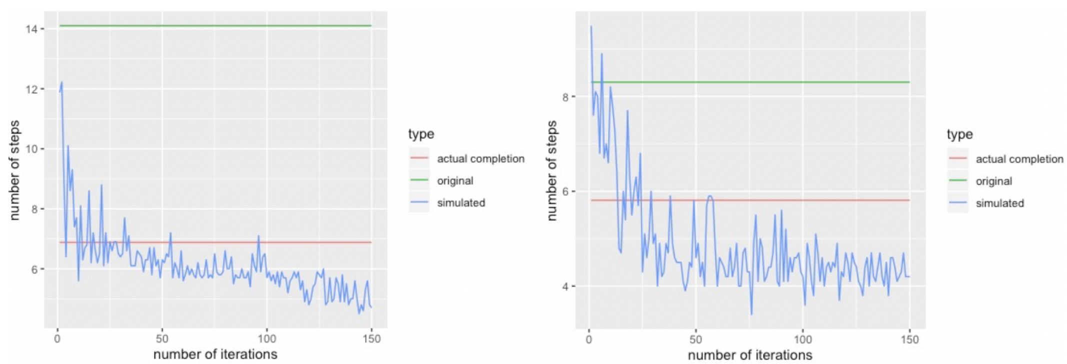


Figure 4.18: Number of steps in INT: Coelacanth (left) and Methane Clathrates (right).

Compared to the two factual tasks (CPE and STP), the two intellectual tasks placed greater challenges on both participants and Q-learning algorithm. As a result, the simulated search sessions achieve relatively smaller improvement in search efficiency and takes more rounds of learning iteration to converge to a better performance (except for the REL-Mathane-Clathrates task where the simulation algorithm significantly improves the performance of search interaction by reducing 3.7 steps on average after the 100th iteration). This may be because these two intellectual amorphous tasks required participants to engage in more high-level cognitive activities (e.g., examine the relationship between two facts, prepare useful materials for different aspects of an interview). Thus, participants were not able to gather all needed information and save useful pages within only one or two queries. This heightened requirement at cognitive level also makes it difficult for our Q-learning algorithm to identify potentially efficient search paths with high rewards in the given solution space and thereby limits the performance of simulated search episodes. In the REL-Coelacanth and two INT tasks, the final improvements achieved in the last two rounds of iterations range from 0.9 to 1.7 query segments. In addition, compared to CPE and STP tasks, in the two intellectual amorphous tasks, it takes more rounds of iterations for the Q-learning algorithm to converge to a range of policies that outperform the actual task completion baseline.

Table 4.9: Comparing search efficiency: average number of steps.

Task-Topic	**Simulated**	**Actual**	**Original**
CPE-C (factual specific)	**2.05****	4.18	8.5
CPE-MC (factual specific)	**2.45****	6.4	12.1
STP-C (factual amorphous)	**2.05***	2.583	3
STP-MC (factual amorphous)	**2.1****	5.727	7.8
REL-C (intellectual amorphous)	**2.21****	3.1	7.9
REL-MC (intellectual amorphous)	**2.3****	5.8	7.6
INT-C (intellectual amorphous)	**4.75****	6.88	14.1
INT-MC (intellectual amorphous)	**4.2****	5.81	8.3

Note: C: Coalacanth; MC: Methane Clathrates; Significant values indicate whether the predictor is significantly better than the best baseline (i.e. actual completion baseline) ($* : p < .05, ** : p < .01$). Statistically significant results are boldfaced.

Table 4.9 compares the results from the last two rounds of Q-learning iterations (20 simulated search sessions for each task-topic combination) and that of the two baselines, namely actual completion session and original search session. The results demonstrate that the Q-learning algorithm built upon our state-based framework significantly outperforms the two baselines to different extents in all task types. The simulated search sessions consisting of recommended query segments from different users reduce the number of steps needed for task completion and improve the efficiency of search interactions. Due to the cognitive challenges behind the two intellectual amorphous tasks, our Q-learning algorithm achieves significant but small improvements compared to the performances in CPE and STP-Methane-Clathrates tasks. This is because there were less high-reward search paths available in associated solution spaces and thus longer search sessions became inevitable in simulated search episodes. Nevertheless, in general, we are still able to support users and improve their search efficiency in REL and INT tasks after more rounds of iterations.

Although the Q-learning algorithm is trained for improving search efficiency (i.e. reducing steps), we are also interested in the performance of simulated search sessions on other dimensions of search interaction. Specifically, to measure the contributions of simulated episodes to users' explorations on the associated topics and on the potential solution space behind the complex task at hand, we add three additional evaluation measures: number of relevant pages visited per query segment, number of unique content pages visited per query segment, and average likelihood of discovery (LD) for each clicked page. The definitions of the two measures are given in the Section 3.5.4. This part of analyses helps us examine and understand the impact of Q-learning-based search simulation in a more comprehensive manner.

4.4.2 Number of relevant pages per query segment

When improving the efficiency of search interaction, does the state-based Q-learning algorithm also influence the exploration aspect of search (in either positive or negative way)? This section aims to partially answer this question by discussing our finding regarding the number of relevant pages visited per query segment in the simulated search episodes generated by Q-learning algorithms. The results are presented in the following figures.

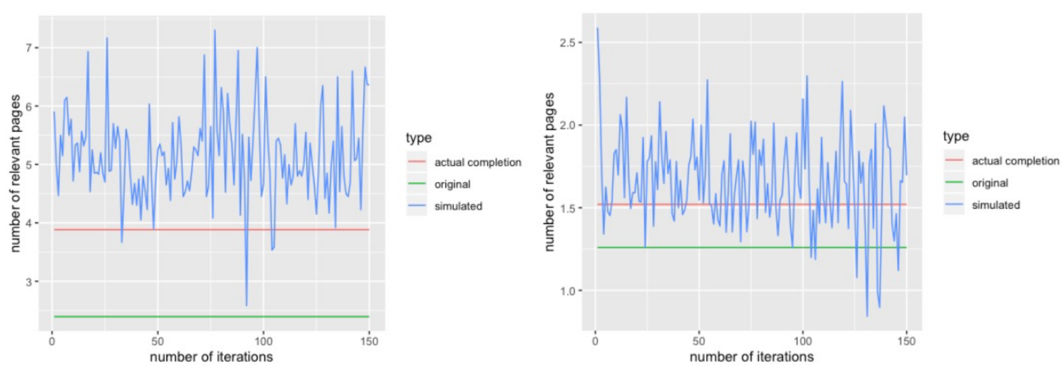


Figure 4.19: Number of relevant pages visited per query segment in CPE: Coelacanth (left) and Methane Clathrates (right).

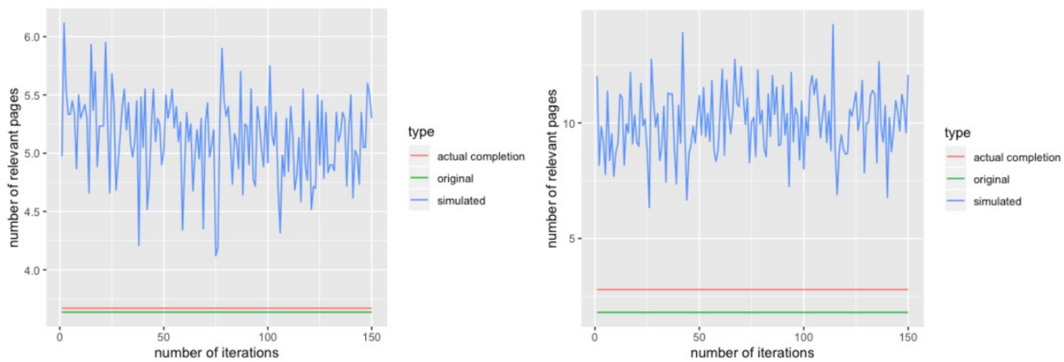


Figure 4.20: Number of relevant pages visited per query segment in STP: Coelacanth (left) and Methane Clathrates (right).

Overall, our result demonstrates that although we do not train the model towards increasing the number of relevant pages, the simulated search episodes do improve the efficiency of finding relevant pages (pages that were clicked by at least two different users) in CPE, STP, and REL tasks. Especially, in the factual amorphous tasks (i.e. STP), compared to the best baseline (i.e. actual completion), the simulated search episodes increase the number of relevant pages visited by nearly 60% on average. This may be because more relevant pages at each step can increase the chance of gathering content pages that are useful for completing the task at hand. As a result, query segments that contain more relevant pages are more frequently captured by the Q-learning algorithms in the iteration processes and thus be included into the simulated search sessions.

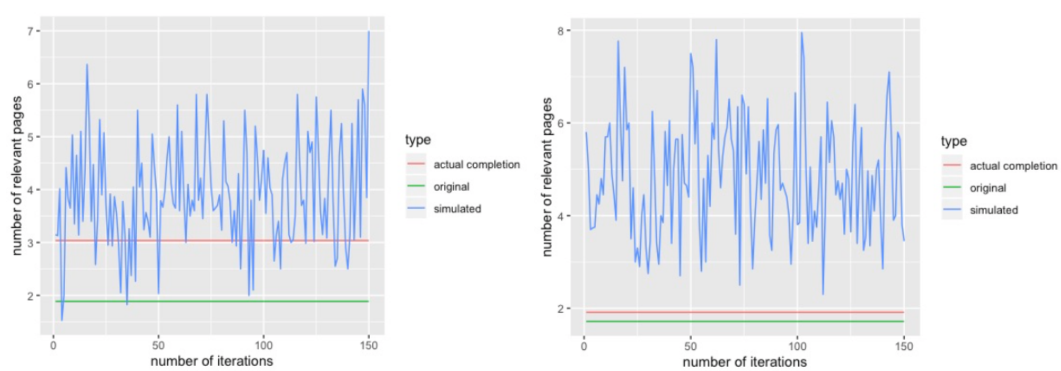


Figure 4.21: Number of relevant pages visited per query segment in REL: Coelacanth (left) and Methane Clathrates (right).

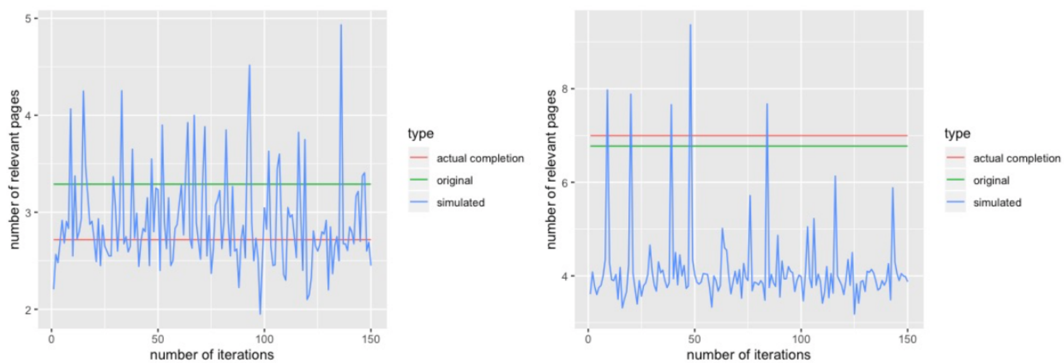


Figure 4.22: Number of relevant pages visited per query segment in INT: Coelacanth (left) and Methane Clathrates (right).

However, according to the results plotted in figure 4.22, this trend does not continue in the INT tasks. Specifically, in the INT-Coelacanth task, there is no significant difference between the simulated search sessions and the actual completion baseline in terms of the number of relevant pages visited per query segment. In the INT-Methane-Clathrates task, there is even a clear decline in the average number of relevant pages clicked. This result indicates that although our Q-learning algorithm improves the efficiency of search by reducing the number of steps needed in INT tasks, it also decreases the number of relevant pages clicked per query segment and may limit the exploration of topically relevant information sources. In other words, the role of Q-learning algorithm here is to foster efficient search by reducing the time spent on the Web pages that are topically relevant but not useful for task completion and helping users reach the focused point of search sooner (cf. [135]). This result also proves that the impact of our state-based Q-learning algorithm on the number of relevant pages per query segment varies across complex search tasks of different types. It is possible that in the INT tasks, the two aspects of search, *finding information that is directly useful for completing the task* and *exploring topically relevant information space*, are not well aligned or even contradict with each other. Consequently, the efforts on supporting one side (i.e. reducing steps needed for task completion may end up "hurting" the other side.

4.4.3 Number of unique pages per query segment

This section presents results about another dimension of search exploration, number of unique pages visited, in the three types of search sessions. This measure focuses on the diversity of pages visited and examines the size of the general information space explored by a user (real or simulated) in each query segment. As is shown in the following plots, in general, the efficiency of search exploration measured by the amount of unique pages visited varies across different tasks and topics. Specifically, In CPE-Mathane-Clathrates task, STP tasks, and REL-Mathane-Clathrates task, the simulated search sessions significantly outperform the corresponding baseline in terms of the number of unique pages visited per query segment. This result indicates that in these search tasks, the Q-learning algorithms not only reduces the steps needed for completing the task, but also improves the efficiency of information space exploration in each step of search. In the CPE-Coelacanth and INT-Coelacanth tasks, the number of unique pages per query segment fluctuates between the average levels of actual completion session and original search session across different rounds of iterations. In these two tasks, the simulated search episodes do not significantly outperform the best baseline, indicating that there is no major improvement from our Q-learning algorithm in the efficiency of exploring broader information space.

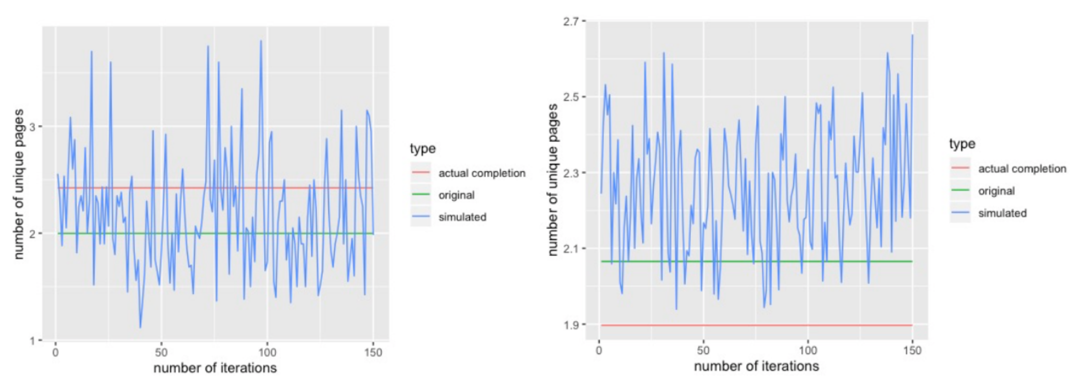


Figure 4.23: Number of unique pages visited per query segment in CPE: Coelacanth (left) and Methane Clathrates (right).

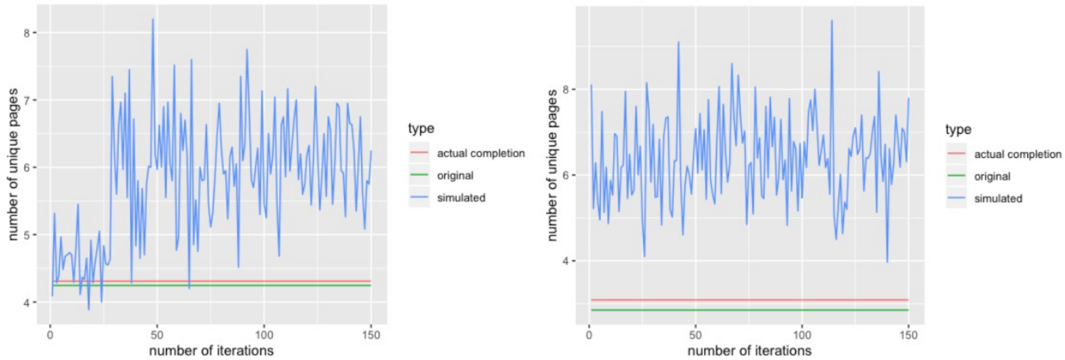


Figure 4.24: Number of unique pages visited per query segment in STP: Coelacanth (left) and Methane Clathrates (right).

In addition, in the REL-Coelacanth and INT-Methane-Clathrates tasks, although the simulated search paths reduce the steps needed for completing the tasks, they also decrease the number of unique pages gathered per query segment and thereby reduce the efficiency of exploring unknown information sources. The divergences in the impact of simulated search episodes on search exploration partially reveal the effects of task type and topic on users' search outcome and tactics in local steps. Also, these divergences suggest that it is necessary to develop different types of search supports for different goals in complex tasks of different types as in some intellectual amorphous search tasks, it is unlikely to simultaneously improve multiple aspects of search interaction (e.g., exploiting useful information, exploring unknown information sources and patches) using one learning and recommendation algorithm.

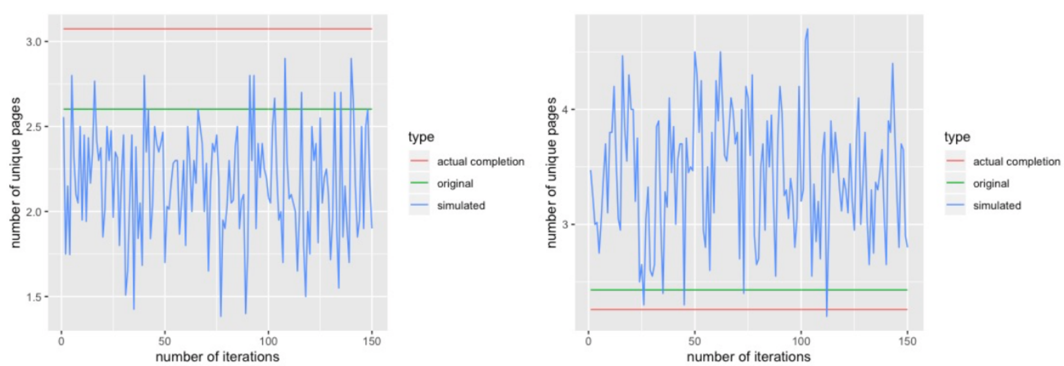


Figure 4.25: Number of unique pages visited per query segment in REL: Coelacanth (left) and Methane Clathrates (right).

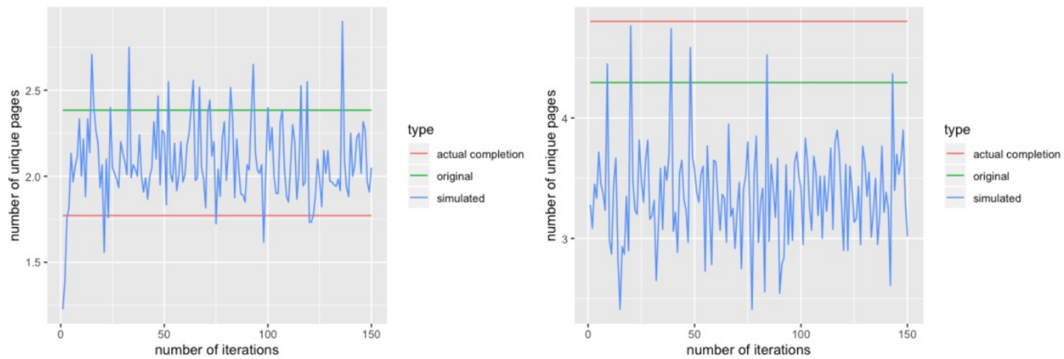


Figure 4.26: Number of unique pages visited per query segment in INT: Coelacanth (left) and Methane Clathrates (right).

4.4.4 Average likelihood of discovery (LD)

Likelihood of discovery (LD) measures the percentage of all page clicks (including repeated visits) a content page received from all users. If a content page is frequently visited by multiple users in many sessions, the page will receive a high LD score. According to [122], the assumption behind this measure is that webpages with a high LD rate are eaiser to find and are common among the majority of users. Thus, these webpages have limited contributions to expanding solution space and exploring unknown information cues and sources. On the other hand, those webpages with a relatively low LD score are difficult to find and may be more useful for exploring new search paths toward completing the search task at hand. Finding low LD webpages can help users discover information that is not just relevant, but also diverse. In this sense, the average LD score allows us to measure the average marginal contribution of each clicked content page to increasing the diversity of discovered information. A content page with lower LD score can be considered as more useful in improving search fairness and diversity.

According to the results presented in the following figures, in the STP-Coelacanth task, the simulated search sessions reach a lower average LD score (compared to that of the two baselines) for most rounds of iterations after the 50th iteration, indicating that in most iterations, the Q-learning algorithm is able to reach a balance between improving search efficiency and increasing the diversity of discovered information (although the algorithm is not specifically trained to find and keep this balance). In other words, compared to pages covered by the actual completion and original search sessions, each visited content page included in the simulated search sessions makes a higher marginal contribution on average to improving session-level search diversity.

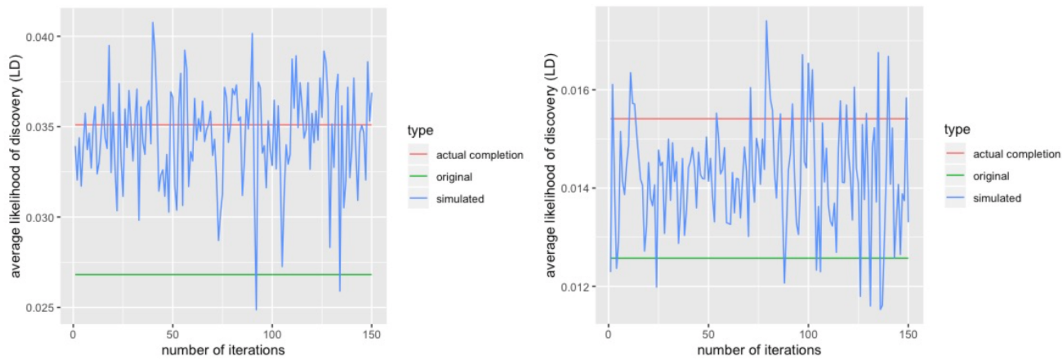


Figure 4.27: Average likelihood of discovery (LD) of each clicked page in CPE: Coelacanth (left) and Methane Clathrates (right).

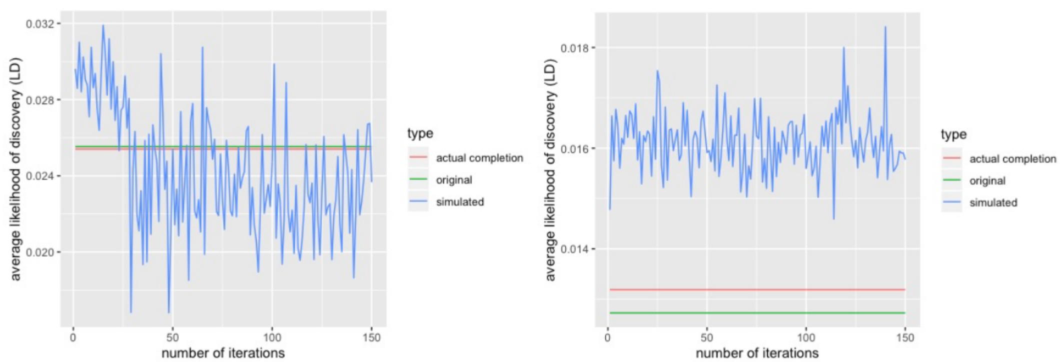


Figure 4.28: Average likelihood of discovery (LD) of each clicked page in STP: Coelacanth (left) and Methane Clathrates (right).

However, in other task types and topics, it tends to be difficult for the algorithm to reach a balance point in the trade-off between improving search efficiency (finding a focused, effective search path) and enhancing the fairness and diversity of discovering information. Specifically, in all other seven task-topic pairs, the average LD scores of each clicked page in simulated search sessions are either clearly higher than or fluctuating between that of the two baselines. The results indicate that in these task-topic combinations, the diversity of discovering information is “sacrificed” for ensuring a good performance in improving search efficiency. When the Q-learning algorithm captures the high-reward, focused search paths, it also reduces the chance for “diverse voice” to come into the simulated search session. For instance, according to the result plotted in the figure 4.16, in the STP-Methane-Clathrates task, as the learning iteration process proceeds, the number of steps needed for task completion drops quickly and gradually converges to a much lower level (between 2 to 2.3 steps on average) compared to the actual completion baseline (5.8 steps). Meanwhile, however, the average LD score of each clicked page in the simulated search sessions is significantly higher than that of the two baseline-type search sessions in all iterations of Q-learning (see figure 4.28). Combining these two part of results together allows us to understand the performance of our state-based framework from two different perspectives: The (simulated) state-based search recommendation may help users reach the focused search stage sooner, but this benefit often comes with the “price” of hurting diversity in information coverage.

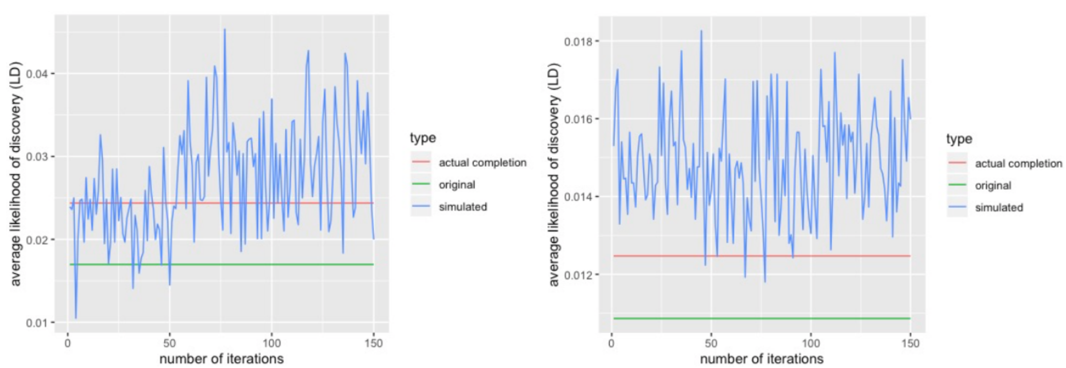


Figure 4.29: Average likelihood of discovery (LD) of each clicked page in REL: Coelacanth (left) and Methane Clathrates (right).

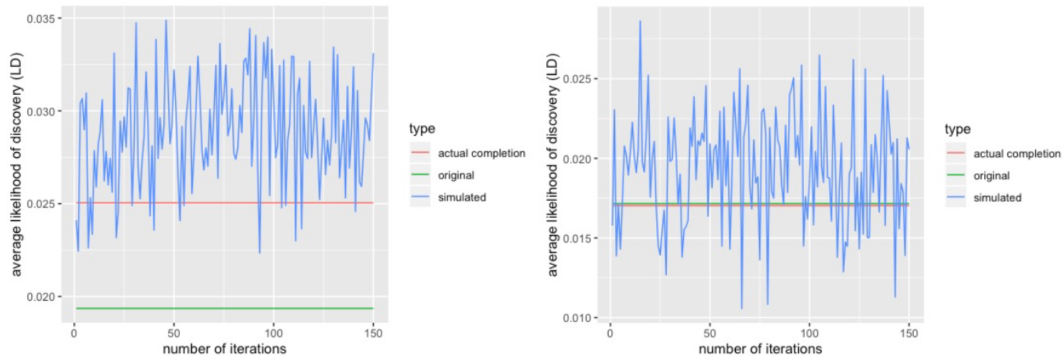


Figure 4.30: Average likelihood of discovery (LD) of each clicked page in INT: Coelacanth (left) and Methane Clathrates (right).

4.5 Summary

This chapter presented in detail the various analyses and simulations conducted on the task-based search datasets collected from user studies. Through our analyses, we identified several intention and problem-help states (RQ1) and explained how the state transition patterns can help us understand and disambiguate complex tasks of different types from a process-oriented perspective (RQ2). Furthermore, we employed search behavioral features to build Machine Learning classifiers and demonstrated that the behavioral-based classifiers can predict the implicit intention and problem-help states with certain levels of accuracy (RQ3). To test the practical value of the proposed state-based framework, we built a Q-learning algorithm based on the knowledge we learned about states and state transition patterns and applied the algorithm in exploring potentially efficient search paths based on the ISI dataset and simulating search path recommendations (RQ4). We then evaluated the simulated search sessions on several dimensions and found a contradiction between search efficiency and diversity of discovering information: Although the simulated search sessions are helpful in reducing the number of steps needed for completing tasks, in many task-topic combinations this improvement on search efficiency often comes with the price of hurting search diversity and hindering the exploration of unknown solution spaces.

Chapter 5

Discussion

In the experiments presented in Chapter 4, we described our results generated from analyzing sequences and transitions of task states and exploring the dynamic nature of complex search tasks from a process-oriented perspective. Most related work so far has focused on the *predefined, static aspects* of complex tasks (e.g., task goal and product [78], task complexity [69], task determinability [27]) and *what* users found as useful information items or objects along the search process rather than *how* the dynamic nature of complex search task is unfolded within the process of search iterations [54, 55]. To explore the dynamic aspect of complex search tasks and understand how predefined task properties are manifested dynamically in search sessions, we used the results of two controlled lab studies and constructed task state framework based on users' information seeking intentions, in-situ search problems and help needed in local search steps. Our approach is unique in 1) understanding planned/intentional and unanticipated aspects of task states and 2) developing a state-based, dynamic approach for recommending search paths extracted from a collective solution space. With respect to the proposed research questions, we have following answers.

5.1 Overall results

To answer the RQ1 (identify task states) and RQ2 (understand state transition patterns), we extracted task states from user annotation data using K-modes clustering algorithms and validated the cluster/state labels via 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. The temporal distributions of different task states also illustrated the differences between complex tasks of varying types from a process-oriented perspective.

The findings related to RQ1 and RQ2 enrich our understanding of the nature of complex search tasks and extend the existing descriptive and computational models of the task-based search process (e.g., [7, 135, 46]) 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 the search task completion process. For instance, in this dissertation study we identified various transition loops and repetitions of task states, and illustrated the differences in nonlinear state transition patterns across different task types. These findings extend Vakkari's model of search process [135] by presenting a higher resolution picture of the recursive processes in task-based search interactions. The state transition loops and repetitions found in our studies could happen in both pre-focus and post-focus stages as it is difficult to identify a clear boundary or break point between pre-focus, exploratory search and post-focus exploitation and known-item searches in a hypothesized linear search process. In addition, the identification of problem-help states and the associated state transition patterns can further enhance our understanding of the factors involved in search processes. Differing from implicit, search behavioral features, extracting and validating states based on users' segment-level annotations enabled us to collect direct empirical evidences on task-related variations at cognitive level and thereby better understand the motivations behind the transitions of search moves and tactics described in previous works (e.g. [7, 126, 53]).

With respect to the RQ3, we investigated the connection between explicit task states extracted from user annotations and implicit search behavioral features (e.g. query formulation, browsing and result examination). Multiple Machine Learning classifiers were built based upon search behavioral measures and the performances of these classifiers in predicting task states were evaluated and compared with two baseline models (i.e. random and most frequent labeling). Our results indicate that both intention-based and problem-help-based task states can be inferred and predicted using behavior-based classifiers with certain levels of accuracy. 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' information seeking intentions and resolving their in-situ search problems at different states of complex search tasks. In addition, the knowledge about task states can be incorporated into user-centered system evaluation processes and facilitate state-based search evaluations. Since users often have different focuses and encounter different obstacles in different task states, a state-aware search system should take this into consideration and focus on different aspects when making recommendations and evaluating search benefits under different states. For example, information coverage and diversity may play an important role in exploratory state, whereas topical relevance metrics (e.g., reciprocal rank, normalized Discounted Cumulative Gain) should receive higher weights when we evaluate system performance under exploitation and known-item task states.

Then, we took a step forward and sought to examine the value of our process-oriented, state-based framework in producing useful search recommendations. Specifically, we built a Q-learning algorithm based on the knowledge we learned about states and state transitions and applied the model in simulating search sessions consisting of potentially useful query segments with high rewards from different users. Q-learning as a Reinforcement Learning method enabled us to run adaptive, step-by-step learning and utilize the knowledge about task states in deciding the strategies of ranking and selecting query segments for recommendation [138]. Our results demonstrate that the simulated search episodes can improve search efficiency to varying extents in complex tasks of different types. In some task contexts (especially intellectual amorphous tasks), however, this improvement often comes with a decrease in the diversity of discovered information and a shrinkage in the exploration of unknown solution space behind task-based search interactions. Given the increasing importance of diversity and fairness in contemporary search and recommender systems [35, 40, 105, 48], more future research efforts are needed on finding and maintaining the balance between developing efficient search supports for complex tasks (e.g., reducing the total amount of steps needed and improving the marginal gain of each step with respect to task completion) and enhancing the diversity, transparency, and fairness of information discovery and interactive search recommendation.

5.2 Implications and limitations

As always, there are lessons learned and broader implications from this dissertation study, limits to our work, as well as needs for future research efforts. A key implication from our work is that Q-learning algorithm as a reinforcement learning method fits well with the problem of developing state-based search supports based on existing solution space or collection of local search paths. Differing from traditional Machine Learning methods, Q-learning algorithm allows systems to learn continuously when new data points keeps coming in and enables IR recommendation systems to iteratively update themselves according to the evolving solution space and the changing benefits (rewards) associated with different ways of search recommendations [138]. Thus, with the recommendation model built upon Q-learning algorithm, we will be able to support users in an online fashion. Nevertheless, it is worth noting that this flexibility is not unconditional. To facilitate the policy update in Q-learning algorithm, we need to have clear definitions and measures regarding the various benefits associated with each unit of action or recommendation (e.g., query segment, search tactic). In this dissertation study, we clearly defined the requirements of search task completion and represented the actual contribution of each bookmarked page using a unique vector. However, it is very difficult, if not entirely impossible, to accurately measure the "benefit" associated with each action in naturalistic search tasks. This is because (1) many search tasks in everyday life and workplace contexts are open-ended in nature and have no clear task completion point (e.g., find useful information about R programming); (2) users' perceived benefits in search sessions are very subjective and are often significantly affected by the gap between remembered utility and experienced utility (cf. [66]); (3) Overall, users' perceived level of search success and actual search performance are not always consistent with each other. Smith and Rieh [124] argued that people often confuse the feeling of being able to find information (through IR systems) with their own actual knowledge learned. A high feeling-of-findability and feeling-of-knowing cannot guarantee a high level of search effectiveness. In addition, given the multidimensionality of search contexts, we need different metrics for measuring different aspects of search gains or benefits (e.g., the efficiency of search task completion, task-specific knowledge learning, overall scope of information exploration). The way in which we measure search benefits fundamentally determines our recommendation approaches as well as system performance evaluations. Using different benefit measures will eventually encourage different types of system supports and push search recommendation into different directions. Given these reasons, designing valid, repeatable measures of search benefits would be a major challenge when we try to generalize the state-based recommendation model to broader contexts.

With respect to limitations, this study only covered 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 [93, 124, 47]). In addition, this study did not measure the accuracy of intention and problem-help annotations. As a result, we are not sure about the extent to which the annotations collected through in-situ and retrospective questionnaires are reliable and repeatable. Nevertheless, this research speaks to a promising direction of conceptually deconstructing and computationally modeling complex search tasks from a process-oriented perspective and may encourage future research to explore the changes and transitions on other dimensions and further enrich the task state framework.

The findings reported here from two small scale user studies need to be tested based on the datasets collected under different topics, task contexts (e.g. structured information search in specialist databases [19, 92]) and study settings (e.g., home environment [37]). To better support users in varying scenarios, we need to expend our study and examine the extent to which a state-based recommendation algorithm developed under a specific topic-task combination could be generalized and "transferred" to a different search context. It is also critical to investigate how user traits (e.g., search skills, cognitive limits, emotional states) and other contextual factors (e.g. other task facets, task relevance to the current situation, network latency) affect the distribution and transition of task states at multiple levels. To address this issue, researchers need to identify and design reliable measures that can capture the nature of these factors in various scenarios and also allow us to differentiate the effect of one factor from that of another contextual factor. For instance, in the context of Web search, it is often difficult to differentiate the impact of topic knowledge from that of topic-independent search skills on search performance. In particular, how to accurately measure a user's level of search skills is still an open question in the area of information retrieval. Instead of merely relying on self-reported metrics, we may be able to develop standard tasks or tests and use users' performances on these tasks (e.g. task completion time, error rate in task performances) as proxies of search skills and topic familiarity measures.

With respect to application and simulation, our state-based Q-learning algorithm was built upon a series of assumptions regarding users' search behaviors and reactions to search recommendations: for instance, we assume that users under the same state have similar search tactics and will receive the same amount of rewards and benefits if given the same query segment. Also, in our simulation, we assume that users always accept the search recommendations provided by our system or algorithm, and that a session with less steps are always more favorable than longer ones. These assumptions allow us to simulate and evaluate search recommendations according to task requirements and temporarily bypass some open HCI problems, such as the individual difference in obtaining and evaluating benefits from recommendations, factors affecting users' acceptance of recommendations across different task contexts (e.g., in-situ relevance of the recommendation, interruptability of the current task, explainability of the recommendation), as well as the sequence and changes of information seeking intentions within individual query segments. Also, in many task-topic combinations, our efficiency-centered algorithm ends up with ignoring the potential benefits from information exploration as these benefits are not included as rewards for training. However, in complex search tasks (especially intellectual amorphous tasks), users may need to do more exploratory search and learning in order to understand the information gathered in previous query segments. Thus, removing these steps from search recommendations may cause obstacles in users' sense making processes. Future research efforts (e.g., user studies in naturalistic settings) are needed for carefully examining the assumptions made for running simulations and adjusting the setup of recommendation experiments based on the findings about real users and authentic problematic situations. For instance, in addition to the number of steps needed for finishing a search session, we also need to take into consideration other forms of costs that may significantly affect users' choices in search interactions, such as the amount of time spent, number of pages visited and perceived cognitive loads. Learning more about users' tasks, knowledge states and cognitive processes may allow us to at least partially release the restrictions associated with our assumptions, revise the (idealized) experimental setup and build more realistic, accurate task state detection and search recommendation models.

It would also be interesting to see how the state-based Q-learning algorithm can be applied in building a real-time search recommendation system and evaluating the performance of search recommendations in a state-based, online fashion (e.g. diversity-oriented evaluation measures in exploration state and efficiency-focused evaluation measures in exploitation state). Matching evaluation metrics with users' actual task states may help us build a stronger connection between the evaluation outcome of search recommendation and users' in-situ search satisfaction. Furthermore, the state-based approach developed in this study could also be applied in other modalities of search interactions, such as mobile search, conversational search, and multi-agent collaborative search. Note that these future applications will also face many of the major challenges discussed above, including the measurement of actual search benefits in different tasks and task states, the divergence between users' perceptions and actual search performances, and the identification of task completion points. In addition to these challenges, we may also encounter problems that are unique to some of the modalities of interaction. For instance, in the context of spoken search interactions, researchers do not have the luxury of collecting data on users' rich interactions with content pages and SERPs. As a result, it is critical to explore and identify new measures (e.g., "answers" from the system, gesture features [52], conversational cues [136]) that can capture the patterns of search interactions and support the training of recommendation algorithms. Besides, we need to take more contextual factors into considerations (e.g., interruptions and cognitive loads of task resumption [95], users' emotional states and prior beliefs [139]) when studying complex search tasks and evaluating search recommendation systems in naturalistic scenarios (e.g., searching on the go [51], reality-based search [23]).

Based on more in-depth user-centered studies in various search contexts and more fine-grained behavioral and evaluation features (e.g., cursor movement signals [58], facial expression features [157], neuro-physiological features [102, 44]) in both model building and system evaluation, it seems realistic to believe that we will eventually be able to develop adaptive search systems that can provide reactive and even proactive task supports for users according to their immediate task states and cognitive abilities.

Chapter 6

Conclusion

6.1 The goals of the dissertation

In this dissertation we sought to 1) understand complex search tasks from a state-based, process-oriented perspective and 2) predict implicit task states from observable behavioral features. Furthermore, through running Q-learning-based simulation of search recommendations, we also aimed to 3) demonstrate how the state-based framework could be applied in building recommendation algorithms and adaptively supporting users engaging in complex search tasks of different types.

6.2 To what extent have the goals been met?

As a general matter, the goals discussed above were largely met.

The fundamental assumption or starting point of this dissertation is that we cannot fully understand the nature of complex search tasks without studying the *process of doing tasks*. Overall, this dissertation presents a comprehensive study on state-based approach to understanding and dynamically supporting complex search tasks, which covers identifying task states (RQ1), examining state transition patterns (RQ2), predicting task states from search behavioral measures (RQ3), and all the way to applying the state-based framework in simulating and evaluating state-based recommendations of query segments (RQ4). Our work connects the theoretical frameworks of search process (both linear and nonlinear) with the computational models of interactive IR and illustrates an innovative analytical approach that is both theoretically meaningful and practically applicable to the anatomy of and support for complex search tasks. Also, the Q-learning-based recommendation algorithm demonstrated in our study can potentially be applied in real-time search recommendations across tasks of varying types.

Our approach can be further extended in other contexts, such as recommendations for multi-session interactive search, multi-user collaborative search, and conversational search. Such further work would enhance the applicability of our state-based framework and shed light on the possible divergence in task state transition patterns and user evaluation of search recommendations across different modalities of search interactions. Also, we could address the limitations associated with the model assumptions in our study by further exploring real users' cognitive processes (especially their cognitive limits and bounded rationality, which are not properly represented in the existing formal models [123, 66]) and strategies of decision-making (e.g., query abandonment, search stopping) in authentic search sessions. This line of future research would further improve our process-oriented dynamic approach to studying complex search tasks and offer our state-based adaptive recommendation model a more solid, realistic psychological and behavioral foundation.

6.3 Broader contributions to interactive IR research

Within the broader context of interactive IR research, this dissertation work pushes the boundary of our knowledge regarding search interactions under complex search tasks (specifically, task states and the nonlinearity of task-based search process). Also, our study helps bridge the gap between descriptive state-based framework and computational model of recommendation simulation and demonstrates a viable way in which the knowledge we learned about users and tasks from small-scale lab studies could be applied in building and evaluating search recommendation systems. In this sense, methodologically, our work may encourage researchers to further explore ways in which the knowledge we learned from user studies of varying types can be better utilized for designing algorithms, making reasonable assumptions, running simulations and eventually building and evaluating adaptive IR systems. The knowledge we learned about real users interacting with information (instead of simulated agents, SERPs, and search sessions) can help us build more accurate, realistic user models and task models and improve the techniques of search personalization and optimization. Following this line of research, we will be able to identify more meaningful research questions and opportunities for understanding people and improving interactive IR systems at the intersection between algorithm-oriented formal modeling and human-centered computing.

6.4 Implications and challenges for practical applications

Going beyond research context, systems that utilize the findings from our dissertation work may generate a series of potential impacts on people's lives in both workspace and everyday life contexts. Meanwhile, when it comes to practical applications, we will also face some extra obstacles in addition to the challenges for IR research explained in the discussion chapter.

Specifically, for instance, the state-based approach acknowledges the variations in users' search tactics, local intentions, and in-situ encountered problems in different task states and provides adaptive supports for users at different points of search interaction. This approach may help search recommendation systems go beyond stateless, ad hoc retrieval paradigm and provides useful supports that can help users accomplish overall goals or tasks behind search sessions (rather than one single query). Users engaging in a complex, broad search task (e.g. learning R programming) can get different types of supports based on their current intention states (e.g., exploring available documentations and programming Q&A communities; searching specifically for information about *ggplot2* package) or encountered problems (e.g. unaware of the available packages for machine learning programming; do not familiar of R-specific terminology; do not know how to express task-related needs with search queries).

With respect to challenges and obstacles, in addition to the challenges discussed in the previous chapter (e.g., measuring benefits in different contexts, understanding the difference between perceived success and actual search performance), one major challenge in real-time state-based recommendation is cold start problem: it will be difficult for an IR system to figure out the task completion criteria before a user actually complete the task. Although we may be able to use the knowledge learned from completed (or abandoned) sessions to estimate the completion points of same or similar search tasks in the future, estimating the completion point for a new task will remain to be a challenge. This will also affect the way we define reward for each step of search and the action of reinforcement learning algorithm. Another possible challenge is that language and topic may limit the generalizability of Q-learning recommendation algorithms. To reduce the potential impacts from these two factors, in this dissertation study we employed behavioral features and benefit measures that are independent from specific topics and languages and can be easily extracted from regular search logs. None of the features was specific to one particular topic or English language. Applying these features in building and training real-time search recommendation algorithms may help us overcome the potential restrictions from language and topic.

6.5 Final thoughts

In summary, the two goals behind the four research questions and the associated analyses are: 1) enhancing our knowledge about the dynamic nature of complex search tasks; 2) leveraging the knowledge in building a computational model that produces adaptive search recommendations and supports users at different points of search. Through achieving these goals, the dissertation also contributes to the integration of the knowledge regarding the dynamic aspect of complex search tasks (i.e. evolving task states) with formal models and computational simulations of search recommendations.

However, it is worth noting that our state-based framework and the associated simulation algorithm are still far from complete in terms of articulating search task characteristics and supporting users dynamically in real time. To better understand how complex search tasks of different types are unfolded in search interactions and (more importantly) why they are manifested in certain ways, we need to overcome the challenges and limitations discussed above and continue our explorations on both human-centered aspect and computational aspect. A more complete development of state-aware, adaptive IR systems would require 1) a more psychologically realistic model of task process built upon deeper knowledge about users' cognitive abilities, knowledge and emotional states, as well as systematic biases, 2) a more comprehensive recommendation algorithm that incorporates more search task features and user traits as parameters, and 3) evaluation metrics that are more closely associated with users' in-situ and whole-session experiences. Besides, with respect to interface and system design, future research also need to explore new affordances and innovative interaction patterns in order to support more sophisticated intentions, task states and the associated search tactics (e.g., summarizing available information about a given topic, selecting the best result from a list of relevant search results, exploring a new domain) that are not well supported by the existing IR systems and query-based interaction paradigm.

Chapter 7

Appendices

7.1 Reward Annotation of Bookmarked Pages: Examples

The Discovery:

Until 1938, coelacanths were known only as an order of peculiar lobe-finned fishes which appeared in the fossil record almost 400 million years ago and then seemed to go extinct about 80 million years ago. So the discovery of a live coelacanth off the coast of South Africa in 1938 was understandably met with great excitement. A subsequent fourteen-year search for a second specimen of this extraordinary fish resulted in the discovery of the "true" home of the living coelacanth, *Latimeria chalumnae*, in the Comoran archipelago in the western Indian Ocean. Since that time about 200 specimens of have been caught in the Comoros. A few other specimens have also been caught near Madagascar and Mozambique, but genetic analyses suggest that these are simply "strays" from the main Comoros population.

The scientific community was shocked again in 1998 when UC Berkeley researchers announced the discovery of a coelacanth in North Sulawesi, Indonesia, almost 10,000 kilometers from the Comoros. Dr. Mark Erdmann first saw a coelacanth in Indonesia in September 1997, while on his honeymoon with his wife, Arnaz. Arriving at a fish market, Arnaz noticed a large, strange-looking fish being wheeled by in a cart. Dr. Erdmann immediately recognized the fish as a coelacanth and excitedly photographed it and briefly interviewed the fisherman. Doubting that they could really have stumbled upon such a major discovery, they concluded that coelacanths must have been found in Indonesia previously. Unfortunately, they did not purchase and preserve the coelacanth. Upon further investigation following their arrival in Berkeley a week later, they found that this was indeed an important and unexpected discovery.

Dr. Erdmann returned to Sulawesi in November, 1997 in search of another coelacanth. During a 5 month period, he interviewed over 200 fisherman in the coastal villages around North Sulawesi, but found very few fisherman who seemed familiar with the fish. Finally, he interviewed two fisherman who said they occasionally caught the coelacanth, which they called *raja laut*, translated as, "king of the sea." After careful monitoring of their catch for several months, Dr. Erdman was rewarded with a second Sulawesi coelacanth on July 30, 1998. The second Sulawesi coelacanth was caught by Om Lameh Sonatham in a deep-set shark gill net off of Manado Tua island in the Bunaken Marine Park (see "where do they live" for map of location). The coelacanth was barely alive when it was delivered to Dr. Erdmann. After they temporarily revived the animal by towing it behind their boat, they photographed it in shallow water (see photo on "home" page). When the injured fish eventually died, it was frozen and later donated to the Indonesian Institute of Sciences. A French and Indonesian research team named the Sulawesi coelacanth as a new species (Pouyand, L., et al. 1999 "A new species of coelacanth." *Comptes Rendus* v322(4):261-267).

In October, 1999, a paper appeared in the Proceedings of the National Academy of Sciences reporting more detailed molecular and morphological analyses of the Sulawesi coelacanth (Holder, M.T., M.V. Erdmann, T.P. Wilcox, R.L. Caldwell and D.M. Hillis. 1999. Two Living Species of Coelacanths? Proc. Nat. Acad. Science 96:12616-12620). Conducted by scientists from the University of Texas, Austin and the University of California at Berkeley, Holder et al. concluded that the Comoran and Sulawesi populations most likely diverged 5.5 million years ago and perhaps as long as 16 million years ago. This is significantly earlier than the 1.2 to 1.4 million years ago suggested by Pouyand et al. The differences are due to sequencing errors made by Pouyand et al. as well as in the way that they calculated the age of divergence. The Texas and Berkeley teams conclude that the molecular evidence suggests that the two populations probably do represent different species. However, when Holder et al. looked at morphological traits reported by Pouyand et al. to differ between the Sulawesi fish and the Comoran population, they did not find them. This is due to the fact that Holder et al. used a larger sample size of Comoran fish which provided a better estimate of the amount of variation that occurs. The definitive answer regarding the relationship of these two populations of *Latimeria* probably won't be determined until additional Indonesian fish are captured and a more detailed morphological study is completed.

The annoucement by Erdmann et al. of the capture and preservation of a living coelacanth almost 10,000 kilometers from the Comoros appeared as a cover story in the journal Nature (Erdmann, M.V., Caldwell, R.L., and Moosa, M.K. 1998. "Indonesian 'King of the Sea' Discovered" Nature v.395:335). The discovery was also featured in television, radio, and newspaper articles around the world; including CNN, ABC News, and National Geographic. Discover magazine even listed the discovery as one of the top science stories of 1998 (January 1999, p.49).

The first formal descriptions of the two known Sulawesi coelacanths and how they were discovered were published in 1999 by Dr. Erdmann in Environmental Biology of Fishes. (Erdmann, M.V. 1999. An account of the first living coelacanth known to scientists in Indonesian waters. Env. Biol. Fishes, 54: 440-444; Erdmann, M.V., R.L Caldwell, S.L. Jewett and A. Tjakrawidjaja. 1999. The second recorded living coelacanth from north Sulawesi. Env. Biology of Fishes, 54: 445-451). In the first paper, Dr. Erdmann describes the events that lead up to the discovery and photographing of the first Indonesian coelacanth in the fish market in Manado. The second paper describes the fish captured off Manado Tua on July 30, 1998, its preservation and donation to the Indonesian government.

Figure 7.1: CPE (Coelacanth) bookmarked page that confirms statements 1 and 2. *Note*: *Statement 1*: The coelacanth ("see-la-kanth") is a 'living fossil' fish, thought to be long extinct before a specimen was discovered in 1938 at East London, South Africa. *Statement 2*: Forty six years after that, a new population was identified from at least two specimens caught off of North Sulawesi, Indonesia. This bookmarked page is represented by the vector $v = \{1, 1, 0, 0, 0, 0\}$ in Q-learning reward measurement.

Home > News

Ancient Human-Size Fish Breathed with Lungs

By Laura Geggel September 15, 2015

At a 427-foot (130 meters) depth off Sodwana Bay in South Africa, the extant coelacanth L. chalumnae swims in its natural environment.
(Image: © Laurent Ballesta / Andromede Oceanology Ltd / http://www.blancpain-ocean-commitment.com/en-us#!/home)

Before the dinosaur age, the coelacanth — a hefty, mysterious fish that now breathes with its gills — sported a well-developed lung, a new study finds.

This lung likely helped the fish survive in low-oxygen, shallow waters hundreds of millions of years ago, the researchers said. During the Mesozoic era, more commonly known as the dinosaur age, it's likely that some species of coelacanth (see-leh-kanth) moved to deeper waters, stopped using their lungs and began relying exclusively on their gills to breathe, the researchers said.

Figure 7.2: CPE (Coelacanth) bookmarked page that confirms the third statement. *Note*: *Statement 3*: Coelacanths are the size of humans. This bookmarked page is represented by the vector $v = \{0, 0, 1, 0, 0, 0\}$ in Q-learning reward measurement.

IARC 2016 Annual Report published

By Carolyn Rosner | *January 30, 2017*

The team of IARC researchers, staff, and students is excited to share this selection of notable activities and achievements with you.

Message from IARC Director Hajo Eicken

Personally, I am honored and happy to serve in my new role as IARC Director, starting in the spring of last year. The transition from my appointment as UAF faculty member was straightforward, thanks to the tireless efforts of my predecessor, Dr.Larry Hinzman. In his tenure as IARC Director 2007–2015, Larry greatly diversified IARC's research portfolio and created a solid foundation for the future. Ultimately, it is the outstanding roster of researchers, staff, and students at IARC that continues to move the institute forward in new directions.

This past year has been particularly interesting, with an unprecedented degree of interest in the Arctic, both at the national and international level. IARC researchers and their findings and insights received significant media attention, and several of us were involved in briefings and panel discussions that were part of the United States' leadership of the Arctic Council and related events, such as the first ever Arctic Science Ministerial Meeting in Washington, D.C. in September 2016.

Arctic opportunities and challenges

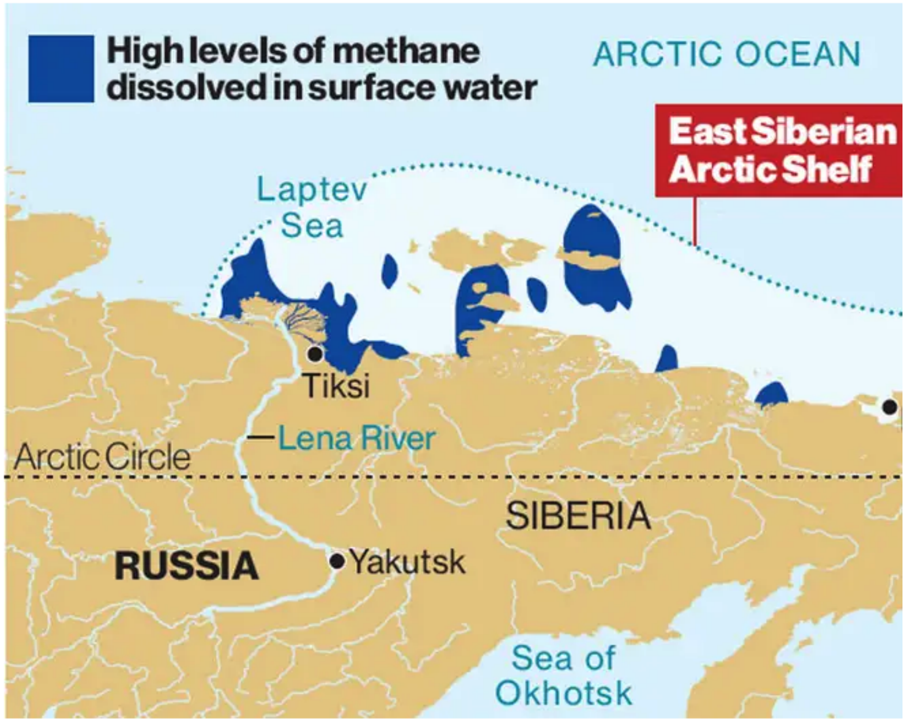
This interest in the Arctic is driven by the recognition that numerous opportunities and challenges are emerging as part of a rapidly changing Arctic. Such changes include economic, geopolitical, and scientific interests in the North from a broad array of non-Arctic nations and organizations.

Thanks to the high caliber and breadth of research, education, outreach, and stakeholder engagement practiced by those collaborating under the IARC umbrella, our institute plays a central role in linking decision-makers to relevant findings and solution-oriented information obtained from research conducted here and elsewhere.

Figure 7.3: INT (Methane Clathrates) bookmarked page that is useful for satisfying the first requirement of the search task.

Note: *The 1st requirement of INT task*: identifying people who likely can provide some personal stories about Dr. Igor Semiletov and his work. This bookmarked page is represented by the vector $v = \{1, 0, 0\}$ in Q-learning reward measurement.

Steve Connor | @SteveAConnor | Tuesday 13 December 2011 11:00 |



Dramatic and unprecedented plumes of methane - a greenhouse gas 20 times more potent than carbon dioxide - have been seen bubbling to the surface of the Arctic Ocean by scientists undertaking an extensive survey of the region.

The scale and volume of the methane release has astonished the head of the Russian research team who has been surveying the seabed of the East Siberian Arctic Shelf off northern Russia for nearly 20 years.

In an exclusive interview with The Independent, Igor Semiletov of the International Arctic Research Centre at the University of Alaska Fairbanks, who led the 8th joint US-Russia cruise of the East Siberian Arctic seas, said that he has never before witnessed the scale and force of the methane being released from beneath the Arctic seabed.

Figure 7.4: INT (Methane Clathrates) bookmarked page that is useful for satisfying the second requirement of the search task.

Note: *The 2nd requirement of INT task*: find the three most interesting facts about Dr. Semiletov's research. This bookmarked page is represented by the vector $v = \{0, 1, 0\}$ in Q-learning reward measurement.

7.2 Intention and Problem-Help Frequency Distributions

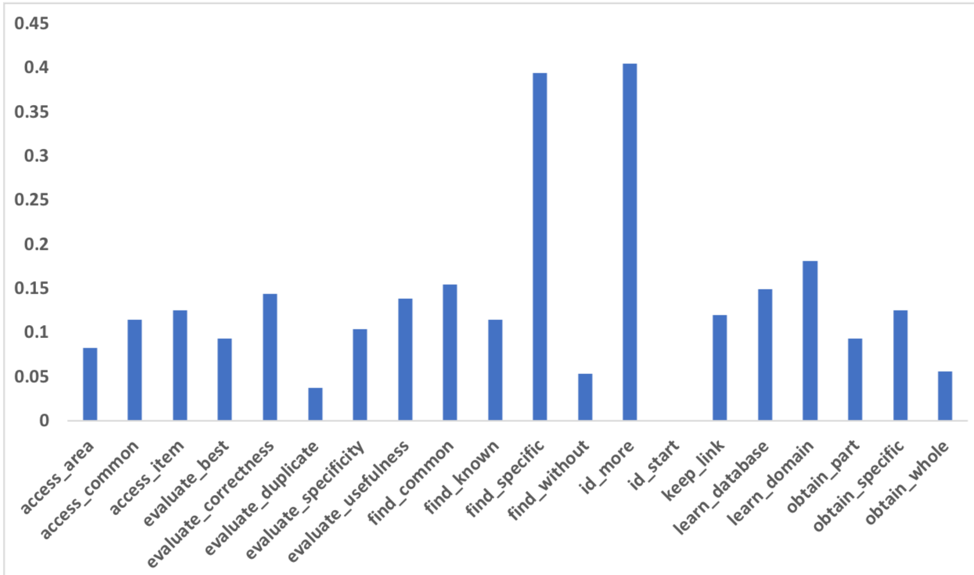


Figure 7.5: Intention frequency distribution: Exploitation.

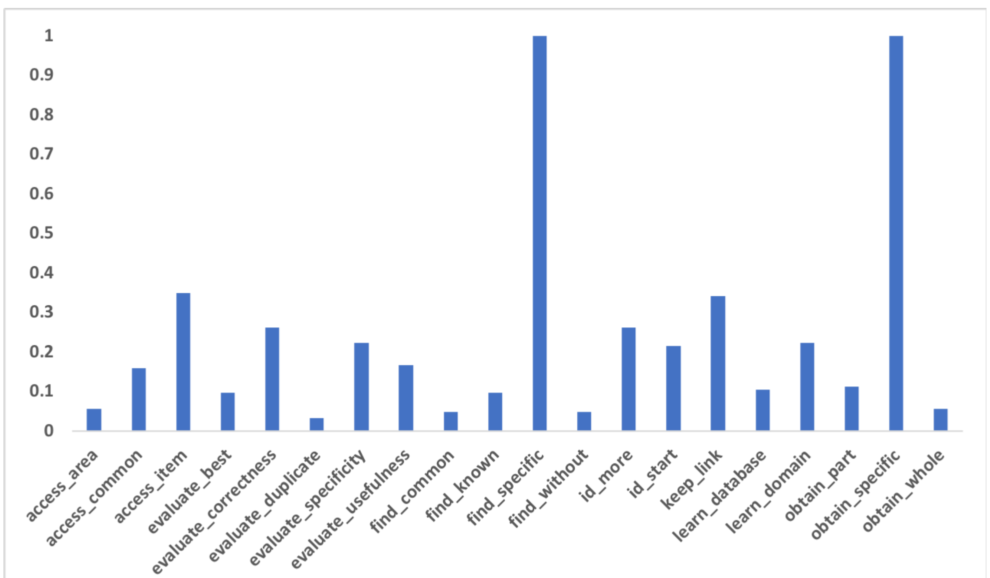


Figure 7.6: Intention frequency distribution: Known-Item.

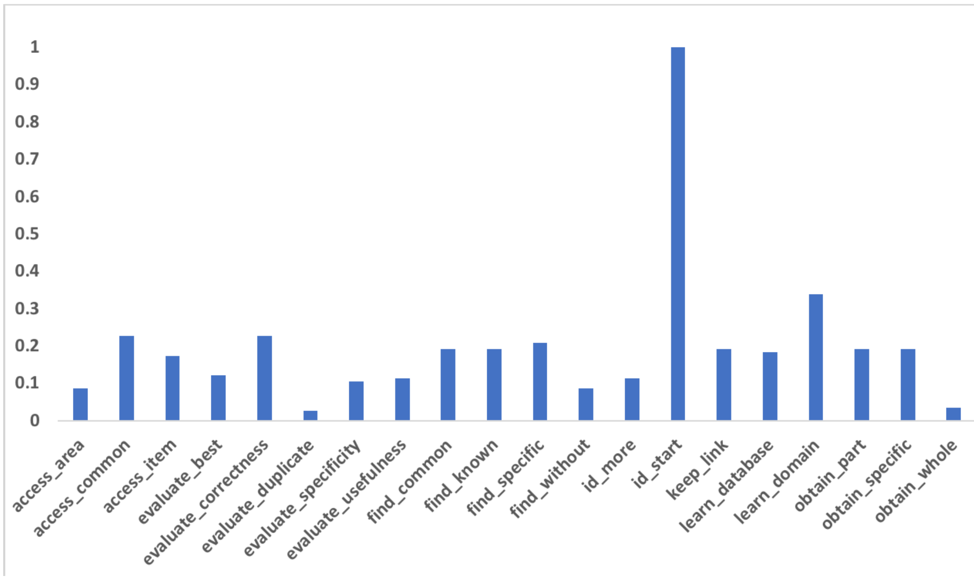


Figure 7.7: Intention frequency distribution: Exploratory.

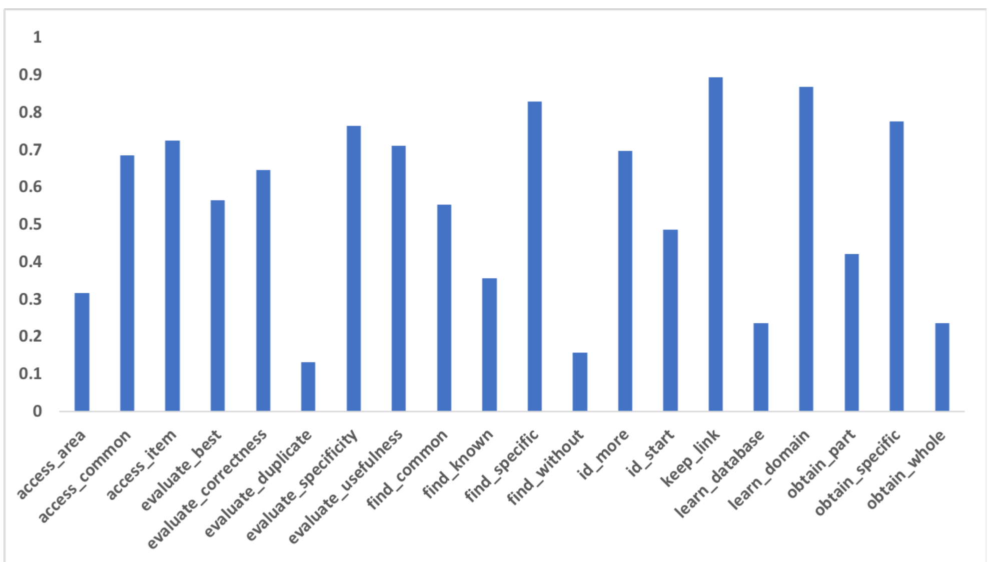


Figure 7.8: Intention frequency distribution: Learning and Evaluation.

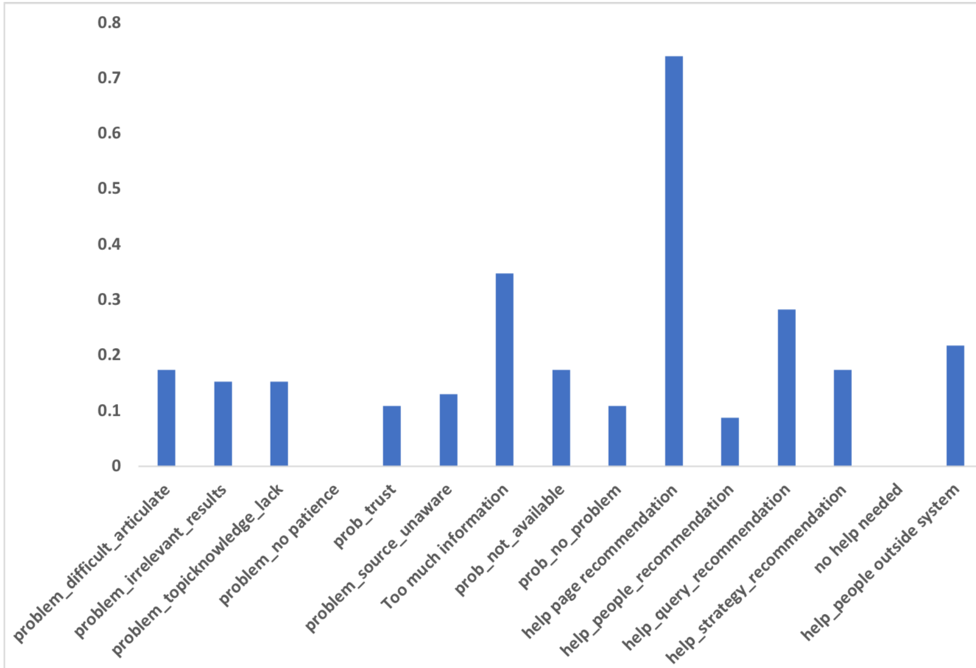


Figure 7.9: Problem-help frequency distribution: IO-P.

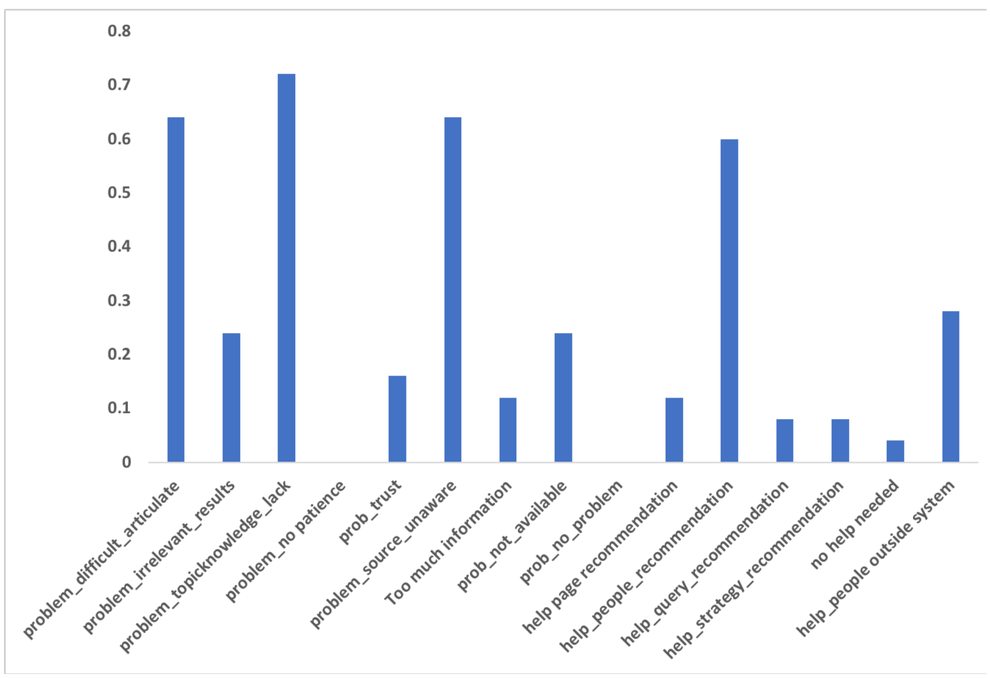


Figure 7.10: Problem-help frequency distribution: ASK-LT-PE.

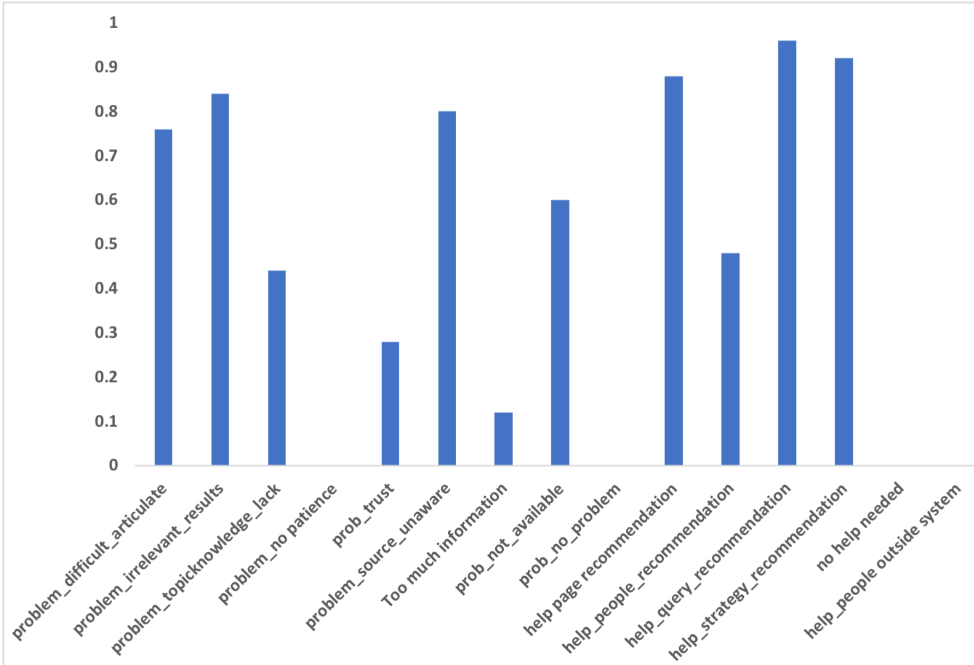


Figure 7.11: Problem-help frequency distribution: ASK-SU-M.

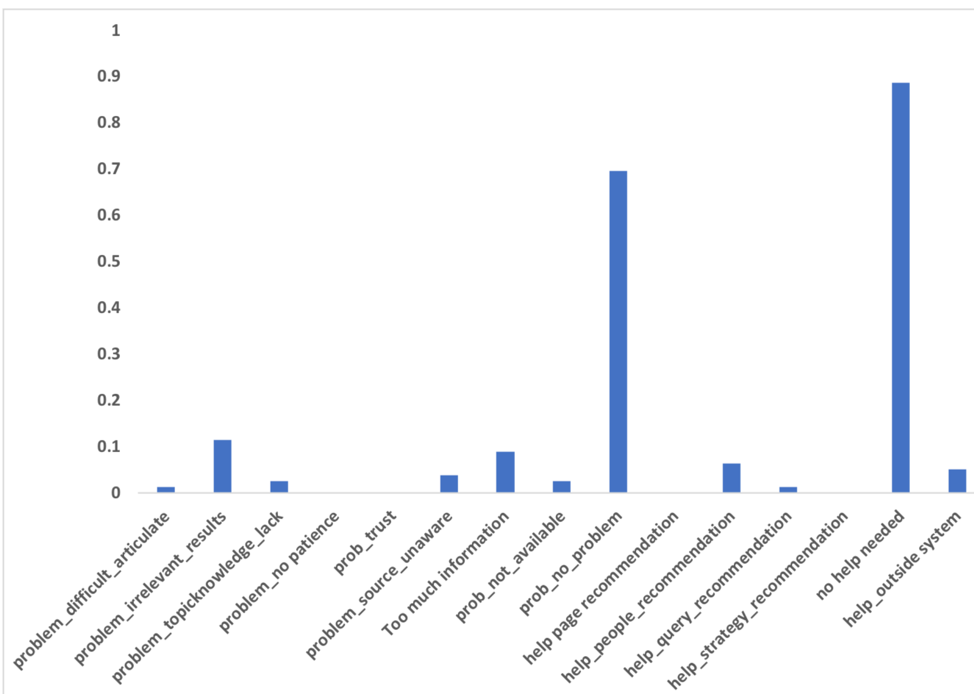


Figure 7.12: Problem-help frequency distribution: NP.

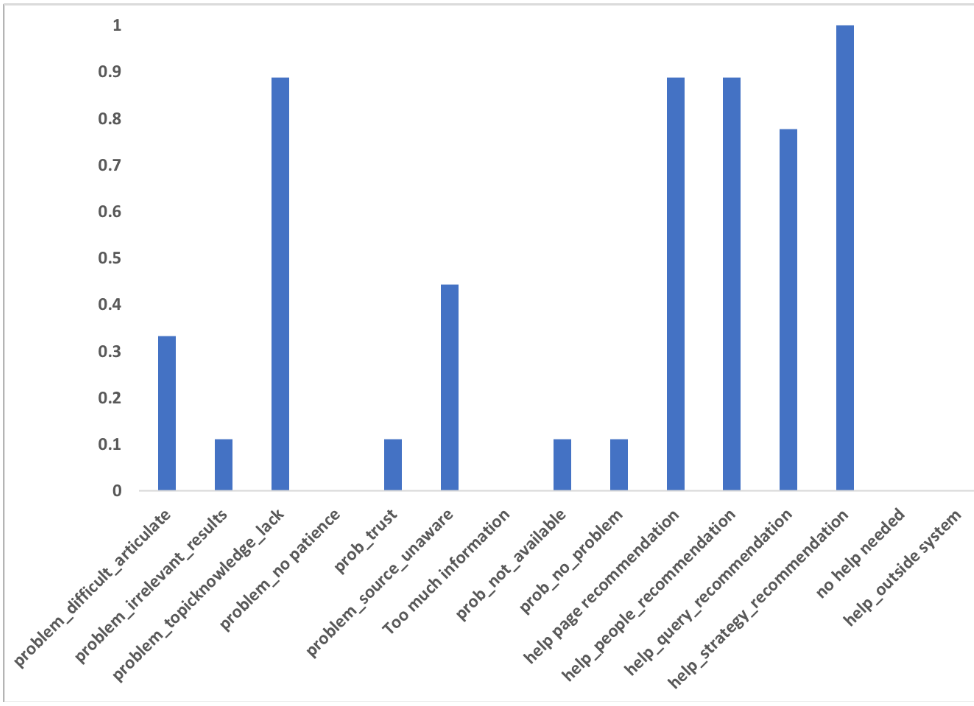


Figure 7.13: Problem-help frequency distribution: LT-M.

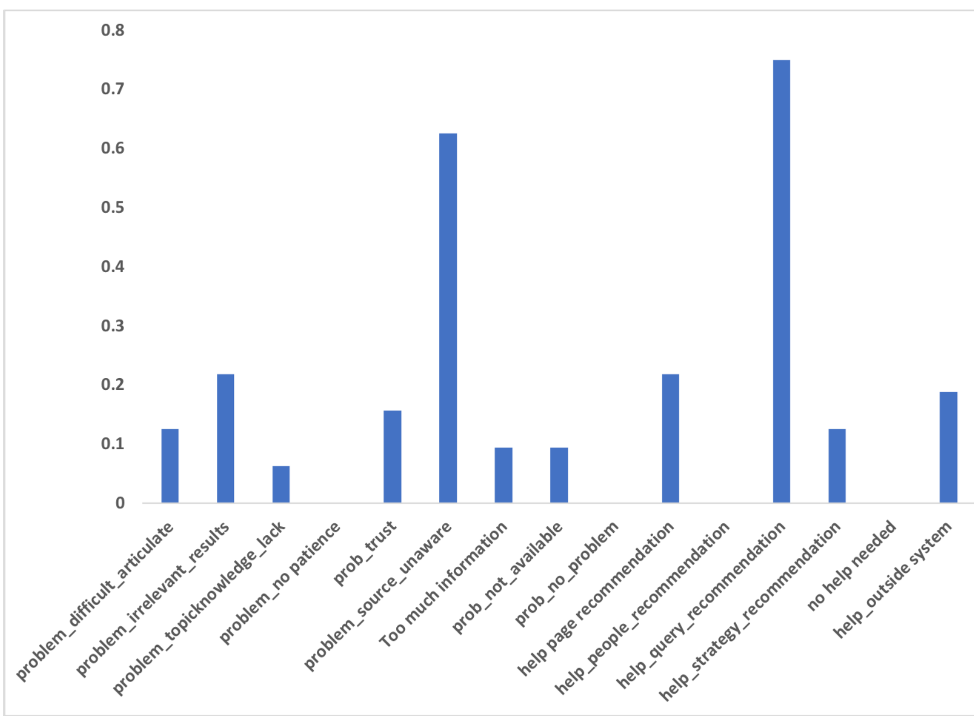


Figure 7.14: Problem-help frequency distribution: SU-QU.

7.3 Results of State Prediction: AUC and F1 scores

Table 7.1: ISI task state prediction: AUC.

Classifier	**Current seg.**	**Prev. session**	**All data**
Logistic Regression	0.679**	0.663**	**0.696****
Support Vector Machine	0.657**	0.642*	0.660**
XGBoost	0.617**	0.622**	0.667**
Random Forest	0.628**	0.619**	0.636**
Decision Tree	0.524	0.517	0.559*
Most Frequent	0.5	0.5	0.5
Random	0.499	0.499	0.499

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

Table 7.2: PH task state prediction: AUC.

Classifier	**Current seg.**	**Prev. session**	**All data**
Logistic Regression	0.677**	0.560	0.681**
Support Vector Machine	0.785**	0.735**	0.776**
XGBoost	0.794**	0.728**	0.783**
Random Forest	**0.803****	0.769**	0.794**
Decision Tree	0.739**	0.712**	0.730**
Most Frequent	0.5	0.5	0.5
Random	0.503	0.503	0.503

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

Table 7.3: PH state type prediction: AUC.

Classifier	**Current seg.**	**Prev. session**	**All data**
Logistic Regression	0.657**	0.483	0.625**
Support Vector Machine	0.801**	0.708**	0.799**
XGBoost	0.813**	0.723**	0.799**
Random Forest	**0.818****	0.747**	0.816**
Decision Tree	0.752**	0.689**	0.745**
Most Frequent	0.5	0.5	0.5
Random	0.501	0.501	0.501

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

Table 7.4: ISI task state prediction: F1 score.

Classifier	**Current seg.**	**Prev. session**	**All data**
Logistic Regression	0.488**	0.479**	**0.518****
Support Vector Machine	0.401	0.466**	0.485**
XGBoost	0.464**	0.481**	0.496**
Random Forest	0.432*	0.477**	0.486**
Decision Tree	0.399	0.407	0.429*
Most Frequent	0.387	0.387	0.387
Random	0.274	0.274	0.274

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

Table 7.5: PH task state prediction: F1 score.

Classifier	**Current seg.**	**Prev. session**	**All data**
Logistic Regression	0.349**	0.2423	0.376**
Support Vector Machine	0.357**	0.314*	0.348**
XGBoost	0.535**	0.522**	0.562**
Random Forest	0.573**	0.546**	0.564**
Decision Tree	**0.583****	0.552**	0.571**
Most Frequent	0.193	0.193	0.193
Random	0.175	0.175	0.175

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

Table 7.6: PH state type prediction: F1 score.

Classifier	**Current seg.**	**Prev. session**	**All data**
Logistic Regression	0.477**	0.431*	0.460*
Support Vector Machine	0.555**	0.454*	0.524**
XGBoost	0.687**	0.522**	0.610**
Random Forest	0.679**	0.621**	0.673**
Decision Tree	**0.694****	0.614**	0.687**
Most Frequent	0.308	0.308	0.308
Random	0.346	0.346	0.346

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

References

[1] Denise E Agosto. Bounded rationality and satisficing in young people's web-based decision making. *Journal of the American society for Information Science and Technology*, 53(1):16–27, 2002.

[2] Denise E Agosto and Sandra Hughes-Hassell. Toward a model of the everyday life information needs of urban teenagers, part 2: Empirical model. *Journal of the American Society for Information Science and Technology*, 57(11):1418–1426, 2006.

[3] Jae-wook Ahn, Peter Brusilovsky, Daqing He, Jonathan Grady, and Qi Li. Personalized web exploration with task models. In *Proceedings of the 17th International Conference on World Wide Web*, pages 1–10. ACM, 2008.

[4] Jaime Arguello. Predicting search task difficulty. In *European Conference on Information Retrieval*, pages 88–99. Springer, 2014.

[5] Anne Aula, Rehan M Khan, and Zhiwei Guan. How does search behavior change as search becomes more difficult? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 35–44. ACM, 2010.

[6] Ahmed Hassan Awadallah, Cathal Gurrin, Mark Sanderson, and Ryen W White. Task intelligence workshop @ WSDM. In *Proceedings of ACM International Conference on Web Search and Data Mining*. ACM, 2019.

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

[8] Jamshid Beheshti, Charles Cole, Dhary Abuhimed, and Isabelle Lamoureux. Tracking middle school students' information behavior via kuhlthau's isp model: Temporality. *Journal of the Association for Information Science and Technology*, 66(5):943–960, 2015.

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

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

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

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

[13] Nicholas J Belkin, Colleen Cool, Adelheit Stein, and Ulrich Thiel. Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems. *Expert systems with applications*, 9(3):379–395, 1995.

[14] Nicholas J. Belkin, Pier Giorgio Marchetti, and Colleen Cool. Braque: Design of an interface to support user interaction in information retrieval. *Information processing & management*, 29(3):325–344, 1993.

[15] Nicholas J Belkin, Robert N Oddy, and Helen M Brooks. Ask for information retrieval: Part i. background and theory. *Journal of documentation*, 38(2):61–71, 1982.

[16] Nicholas J Belkin, Thomas Seeger, and Gernot Wersig. Distributed expert problem treatment as a model for information system analysis and design. *Journal of Information Science*, 5(5):153–167, 1982.

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

[18] Robert McCallum Blumenthal and Ronald Kay Getoor. *Markov processes and potential theory*. Courier Corporation, 2007.

[19] Dario Bonino, Alberto Ciaramella, and Fulvio 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.

[20] Pia Borlund. The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. *Information Research. An International Electronic Journal*, 8(3), 2003.

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

[22] Robin Burke. Knowledge-based recommender systems. *Encyclopedia of library and information systems*, 69(Supplement 32):175–186, 2000.

[23] Wolfgang Büschel, Annett Mitschick, and Raimund Dachselt. Here and now: Reality-based information retrieval: Perspective paper. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, pages 171–180, 2018.

[24] Katriina Byström and Preben Hansen. Work tasks as units for analysis in information seeking and retrieval studies. *Emerging frameworks and methods*, pages 239–251, 2002.

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

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

[27] Robert Capra, Jaime Arguello, Heather O'Brien, Yuan Li, and Bogeum 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.

[28] Robert Capra, Jaime Arguello, and Yinglong Zhang. The effects of search task determinability on search behavior. In *European Conference on Information Retrieval*, pages 108–121. Springer, 2017.

[29] Anil Chaturvedi, Paul E Green, and J Douglas Caroll. K-modes clustering. *Journal of Classification*, 18(1):35–55, 2001.

[30] Ye Chen, Ke Zhou, Yiqun Liu, Min Zhang, and Shaoping Ma. Meta-evaluation of online and offline web search evaluation metrics. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 15–24. ACM, 2017.

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

[32] Jacob Cohen. *Statistical power analysis for the behavioral sciences.* Routledge, 2013.

[33] Michael J Cole, Jacek Gwizdka, Ralf Bierig, Nicholas J Belkin, Jingjing Liu, Chang Liu, and Xiangmin Zhang. Linking search tasks with low-level eye movement patterns. In *Proceedings of the 28th Annual European Conference on Cognitive Ergonomics*, pages 109–116. ACM, 2010.

[34] Michael J Cole, Chathra Hendahewa, Nicholas J Belkin, and Chirag Shah. User activity patterns during information search. *ACM Transactions on Information Systems (TOIS)*, 33(1):1, 2015.

[35] J Shane Culpepper, Fernando Diaz, and Mark 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.

[36] Brenda Dervin. Sense-making theory and practice: an overview of user interests in knowledge seeking and use. *Journal of knowledge management*, 2(2):36–46, 1998.

[37] Allison Druin, Elizabeth Foss, Hilary Hutchinson, Evan Golub, and Leshell 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.

[38] Sebastian Dungs and Norbert Fuhr. Advanced hidden markov models for recognizing search phases. In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*, pages 257–260. ACM, 2017.

[39] Ashlee Edwards and Diane Kelly. Engaged or frustrated?: Disambiguating emotional state in search. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 125–134. ACM, 2017.

[40] Michael D Ekstrand, Robin Burke, and Fernando Diaz. Fairness and discrimination in retrieval and recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1403–1404, 2019.

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

[42] David Ellis. Modeling the information-seeking patterns of academic researchers: A grounded theory approach. *The Library Quarterly*, 63(4):469–486, 1993.

[43] David Ellis, Deborah Cox, and Katherine Hall. A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of documentation*, 49(4):356–369, 1993.

[44] Manuel JA Eugster, Tuukka Ruotsalo, Michiel M Spapé, Ilkka Kosunen, Oswald Barral, Niklas Ravaja, Giulio Jacucci, and Samuel Kaski. Predicting termrelevance from brain signals. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 425–434, 2014.

[45] Henry A Feild, James Allan, and Rosie 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.

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

[47] Ujwal Gadiraju, Ran Yu, Stefan Dietze, and Peter Holtz. Analyzing knowledge gain of users in informational search sessions on the web. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, pages 2–11, 2018.

[48] Ruoyuan Gao and Chirag Shah. How fair can we go: Detecting the boundaries of fairness optimization in information retrieval. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval*, pages 229–236, 2019.

[49] Souvick Ghosh, Manasa Rath, and Chirag Shah. Searching as learning: Exploring search behavior and learning outcomes in learning-related tasks. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, pages 22–31. ACM, 2018.

[50] Jacek Gwizdka. Distribution of cognitive load in web search. *Journal of the American Society for Information Science and Technology*, 61(11):2167–2187, 2010.

[51] Morgan Harvey and Matthew Pointon. Searching on the go: the effects of fragmented attention on mobile web search tasks. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 155–164, 2017.

[52] Larry Heck, Dilek Hakkani-Tür, Madhu Chinthakunta, Gokhan Tur, Rukmini Iyer, Partha Parthasarathy, Lisa Stifelman, Elizabeth Shriberg, and Ashley Fidler. Multi-modal conversational search and browse. In *First Workshop on Speech, Language and Audio in Multimedia*, 2013.

[53] Chathra Hendahewa and Chirag Shah. Segmental analysis and evaluation of user focused search process. In *2013 12th International Conference on Machine Learning and Applications*, volume 1, pages 291–294. IEEE, 2013.

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

[55] Chathra Hendahewa and Chirag Shah. Evaluating user search trails in exploratory search tasks. *Information Processing & Management*, 53(4):905–922, 2017.

[56] Jonathan L Herlocker, Joseph A Konstan, Al Borchers, and John Riedl. An algorithmic framework for performing collaborative filtering. In *22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 1999*, pages 230–237. Association for Computing Machinery, Inc, 1999.

[57] Ronald A Howard. Dynamic programming and markov processes. 1960.

[58] Jeff Huang, Ryen W White, and Susan 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.

[59] Hugo Christian Huurdeman. Dynamic compositions: Recombining search user interface features for supporting complex work tasks. In *CEUR Workshop Proceedings*, volume 1798, pages 22–25. Technical University of Aachen, 2017.

[60] Peter Ingwersen. Cognitive perspectives of information retrieval interaction: elements of a cognitive ir theory. *Journal of documentation*, 52(1):3–50, 1996.

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

[62] Bernard J Jansen and Danielle 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.

[63] Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446, 2002.

[64] Jiepu Jiang, Ahmed Hassan Awadallah, Xiaolin Shi, and Ryen 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.

[65] Jiepu Jiang, Daqing He, and James Allan. Searching, browsing, and clicking in a search session: changes in user behavior by task and over time. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 607–616. ACM, 2014.

[66] Daniel Kahneman. Maps of bounded rationality: Psychology for behavioral economics. *American economic review*, 93(5):1449–1475, 2003.

[67] Junzo Kamahara, Tomofumi Asakawa, Shinji Shimojo, and Hideo Miyahara. A community-based recommendation system to reveal unexpected interests. In *11th international multimedia modelling conference*, pages 433–438. IEEE, 2005.

[68] Melanie Kellar, Carolyn Watters, and Michael 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.

[69] Diane Kelly, Jaime Arguello, Ashlee Edwards, and Wan-ching 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.

[70] Diane Kelly and Nicholas J Belkin. Display time as implicit feedback: understanding task effects. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 377–384. ACM, 2004.

[71] Marijn Koolen, Jaap Kamps, Toine Bogers, Nicholas Belkin, Diane Kelly, and Emine Yilmaz. Report on the second workshop on supporting complex search tasks. In *ACM SIGIR Forum*, volume 51, pages 58–66. ACM, 2017.

[72] Carol 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.

[73] Carol C Kuhlthau. A principle of uncertainty for information seeking. *Journal of documentation*, 49(4):339–355, 1993.

[74] Carol C Kuhlthau, Jannica Heinström, and Ross J Todd. The information search process revisited: Is the model still useful. *Information research*, 13(4):13–4, 2008.

[75] Sanna Kumpulainen and Kalervo Järvelin. Barriers to task-based information access in molecular medicine. *Journal of the American Society for Information Science and Technology*, 63(1):86–97, 2012.

[76] J Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977.

[77] Yuelin Li. Exploring the relationships between work task and search task in information search. *Journal of the American Society for information Science and Technology*, 60(2):275–291, 2009.

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

[79] Yuelin Li and Nicholas J Belkin. An exploration of the relationships between work task and interactive information search behavior. *Journal of the American Society for information Science and Technology*, 61(9):1771–1789, 2010.

[80] Shin-Jeng Lin and Nicholas J Belkin. Validation of a model of information seeking over multiple search sessions. *Journal of the American Society for Information Science and Technology*, 56(4):393–415, 2005.

[81] Chang Liu, Nicholas J Belkin, and Michael J Cole. Personalization of search results using interaction behaviors in search sessions. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 205–214. ACM, 2012.

[82] Chang Liu, Jingjing Liu, and Nicholas J Belkin. Predicting search task difficulty at different search stages. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 569–578. ACM, 2014.

[83] Chang Liu, Ying-Hsang Liu, Tom Gedeon, Yu Zhao, Yiming Wei, and Fan Yang. The effects of perceived chronic pressure and time constraint on information search behaviors and experience. *Information Processing & Management*, 56(5):1667–1679, 2019.

[84] Jingjing Liu, Michael J Cole, Chang Liu, Ralf Bierig, Jacek Gwizdka, Nicholas J Belkin, Jun Zhang, and Xiangmin Zhang. Search behaviors in different task types. In *Proceedings of the 10th Annual Joint Conference on Digital Libraries*, pages 69–78. ACM, 2010.

[85] Jingjing Liu, Jacek Gwizdka, Chang Liu, and Nicholas J Belkin. Predicting task difficulty for different task types. In *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem-Volume 47*, page 16. American Society for Information Science, 2010.

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

[87] Jiqun Liu. A reference-dependent model of search evaluation. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, pages 405–408. ACM, 2019.

[88] Jiqun Liu, Matthew Mitsui, Nicholas J Belkin, and Chirag 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.

[89] Jiqun Liu and Chirag Shah. Investigating the impacts of expectation disconfirmation on web search. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, pages 319–323. ACM, 2019.

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

[91] Jiqun Liu, Yiwei Wang, Soumik Mandal, and Chirag Shah. Exploring the immediate and short-term effects of peer advice and cognitive authority on web search behavior. *Information Processing & Management*, 56(3):1010–1025, 2019.

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

[93] Irene 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.

[94] Jiyun Luo, Sicong Zhang, and Hui Yang. Win-win search: Dual-agent stochastic game in session search. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 587–596. ACM, 2014.

[95] Gloria Mark, Daniela Gudith, and Ulrich Klocke. The cost of interrupted work: more speed and stress. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 107–110, 2008.

[96] Rishabh Mehrotra and Emine Yilmaz. Extracting hierarchies of search tasks & subtasks via a bayesian nonparametric approach. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 285–294. ACM, 2017.

[97] Francisco S Melo. Convergence of q-learning: A simple proof. *Institute Of Systems and Robotics, Tech. Rep*, pages 1–4, 2001.

[98] Alessandro Micarelli, Fabio Gasparetti, Filippo Sciarrone, and Susan Gauch. Personalized search on the world wide web. In *The adaptive web*, pages 195–230. Springer, 2007.

[99] Tomas Mikolov, Edouard Grave, Piotr Bojanowski, Christian Puhrsch, and Armand Joulin. Advances in pre-training distributed word representations. *arXiv preprint arXiv:1712.09405*, 2017.

[100] Matthew Mitsui, Jiqun Liu, Nicholas J Belkin, and Chirag 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.

[101] Matthew Mitsui, Chirag Shah, and Nicholas J Belkin. Extracting information seeking intentions for web search sessions. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 841–844. ACM, 2016.

[102] Yashar Moshfeghi, Peter Triantafillou, and Frank 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.

[103] Allen Newell and Herbert Alexander Simon. *Human problem solving*, volume 104. Prentice-Hall Englewood Cliffs, NJ, 1972.

[104] Robert N Oddy. Information retrieval through man-machine dialogue. *Journal of documentation*, 33(1):1–14, 1977.

[105] Alexandra Olteanu, Jean Garcia-Gathright, Maarten de Rijke, and Michael D Ekstrand. Workshop on fairness, accountability, confidentiality, transparency, and safety in information retrieval (facts-ir). In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1423–1425, 2019.

[106] Michael J Pazzani and Daniel Billsus. Content-based recommendation systems. In *The adaptive web*, pages 325–341. Springer, 2007.

[107] George E Raptis, Christos Fidas, Christina Katsini, and Nikolaos Avouris. A cognition-centered personalization framework for cultural-heritage content. *User Modeling and User-Adapted Interaction*, 29(1):9–65, 2019.

[108] Eun Youp Rha, Matthew Mitsui, Nicholas J Belkin, and Chirag Shah. Exploring the relationships between search intentions and query reformulations. *Proceedings of the Association for Information Science and Technology*, 53(1):1–9, 2016.

[109] Eun Youp Rha, Wei Shi, and Nicholas J Belkin. An exploration of reasons for query reformulations. *Proceedings of the Association for Information Science and Technology*, 54(1):337–346, 2017.

[110] Francesco Ricci, Lior Rokach, and Bracha Shapira. Introduction to recommender systems handbook. In *Recommender systems handbook*, pages 1–35. Springer, 2011.

[111] Xin Rong. word2vec parameter learning explained. *arXiv preprint arXiv:1411.2738*, 2014.

[112] Daniel E Rose and Danny Levinson. Understanding user goals in web search. In *Proceedings of the 13th international conference on World Wide Web*, pages 13–19. ACM, 2004.

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

[114] Laila Safoury and Akram Salah. Exploiting user demographic attributes for solving cold-start problem in recommender system. *Lecture Notes on Software Engineering*, 1(3):303–307, 2013.

[115] George Samaras. Utilizing human cognitive and emotional factors for user-centered computing. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*, pages 191–192, 2017.

[116] Shawon Sarkar, Matthew Mitsui, Jiqun Liu, and Chirag Shah. Implicit information need as explicit problems, help, and behavioral signals. *Information Processing & Management*, 57:102069, 2020.

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

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

[119] Reijo Savolainen. Approaches to socio-cultural barriers to information seeking. *Library & information science research*, 38(1):52–59, 2016.

[120] Reijo Savolainen and Jarkko Kari. Facing and bridging gaps in web searching. *Information Processing & Management*, 42(2):519–537, 2006.

[121] Chirag Shah. Information fostering-being proactive with information seeking and retrieval: Perspective paper. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, pages 62–71. ACM, 2018.

[122] Chirag Shah and Roberto González-Ibáñez. Evaluating the synergic effect of collaboration in information seeking. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 913–922. ACM, 2011.

[123] Herbert A Simon. Bounded rationality in social science: Today and tomorrow. *Mind & Society*, 1(1):25–39, 2000.

[124] Catherine L Smith and Soo Young 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.

[125] Maximilian Speicher, Andreas Both, and Martin Gaedke. Sos: Does your search engine results page (serp) need help? In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1005–1014. ACM, 2015.

[126] Amanda Spink. Study of interactive feedback during mediated information retrieval. *Journal of the american society for information science*, 48(5):382–394, 1997.

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

[128] Lucy A Suchman. *Plans and situated actions: The problem of human-machine communication.* Cambridge University Press, 1987.

[129] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction.* MIT press, 2018.

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

[131] Jaime Teevan, Christine Alvarado, Mark S Ackerman, and David R Karger. The perfect search engine is not enough: a study of orienteering behavior in directed search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 415–422. ACM, 2004.

[132] Jaime Teevan, Susan T Dumais, and Eric Horvitz. Characterizing the value of personalizing search. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 757–758. ACM, 2007.

[133] Jaime Teevan, Susan T Dumais, and Eric Horvitz. Potential for personalization. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 17(1):4, 2010.

[134] Vu Tran, David Maxwell, Norbert Fuhr, and Leif Azzopardi. Personalised search time prediction using markov chains. In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*, pages 237–240. ACM, 2017.

[135] Pertti 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.

[136] Alexandra Vtyurina and Adam Fourney. Exploring the role of conversational cues in guided task support with virtual assistants. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 208. ACM, 2018.

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

[138] Christopher Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.

[139] Ryen White. Beliefs and biases in web search. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 3–12, 2013.

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

[141] Ryen W White. *Interactions with search systems.* Cambridge University Press, 2016.

[142] Ryen W White and Ahmed Hassan Awadallah. Task duration estimation. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 636–644. ACM, 2019.

[143] Ryen W White and Diane Kelly. A study on the effects of personalization and task information on implicit feedback performance. In *Proceedings of the 15th ACM international conference on Information and knowledge management*, pages 297–306. ACM, 2006.

[144] Barbara M Wildemuth. The effects of domain knowledge on search tactic formulation. *Journal of the american society for information science and technology*, 55(3):246–258, 2004.

[145] Thomas D Wilson. Human information behavior. *Informing science*, 3(2):49–56, 2000.

[146] Tom D Wilson. Models in information behaviour research. *Journal of documentation*, 55(3):249–270, 1999.

[147] Wan-Ching Wu and Diane Kelly. Online search stopping behaviors: An investigation of query abandonment and task stopping. *Proceedings of the American Society for Information Science and Technology*, 51(1):1–10, 2014.

[148] Wan-Ching Wu, Diane Kelly, Ashlee Edwards, and Jaime Arguello. Grannies, tanning beds, tattoos and nascar: Evaluation of search tasks with varying levels of cognitive complexity. In *Proceedings of the 4th Information Interaction in Context Symposium*, pages 254–257. ACM, 2012.

[149] Michael Wunder, Michael L Littman, and Monica Babes. Classes of multiagent q-learning dynamics with epsilon-greedy exploration. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 1167–1174. Citeseer, 2010.

[150] Hong Xie. Patterns between interactive intentions and information-seeking strategies. *Information Processing and Management*, 38(1):55–77, 2002.

[151] Iris Xie. *Interactive information retrieval in digital environments.* IGI Global, 2008.

[152] Iris Xie and Colleen 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.

[153] Songhua Xu, Hao Jiang, and Francis Lau. User-oriented document summarization through vision-based eye-tracking. In *Proceedings of the 14th international conference on Intelligent user interfaces*, pages 7–16. ACM, 2009.

[154] Grace Hui Yang, Xuchu Dong, Jiyun Luo, and Sicong Zhang. Session search modeling by partially observable markov decision process. *Information Retrieval Journal*, 21(1):56–80, 2018.

[155] Hema Yoganarasimhan. Search personalization using machine learning. *Management Science*, 2019.

[156] Xiaojun Yuan and Nicholas J Belkin. Supporting multiple information-seeking strategies in a single system framework. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 247–254. ACM, 2007.

[157] Marzieh Yari Zanganeh and Nadjla Hariri. The role of emotional aspects in the information retrieval from the web. *Online Information Review*, 2018.

[158] Pengyi Zhang and Dagobert Soergel. Towards a comprehensive model of the cognitive process and mechanisms of individual sensemaking. *Journal of the Association for Information Science and Technology*, 65(9):1733–1756, 2014.

[159] Pengyi Zhang and Dagobert Soergel. Process patterns and conceptual changes in knowledge representations during information seeking and sensemaking: A qualitative user study. *Journal of Information Science*, 42(1):59–78, 2016.

[160] Xiangmin Zhang, Jingjing Liu, Michael Cole, and Nicholas Belkin. Predicting users' domain knowledge in information retrieval using multiple regression analysis of search behaviors. *Journal of the Association for Information Science and Technology*, 66(5):980–1000, 2015.