An Integrated Model of Tasks and Uncertainties for Designing Task-aware Search Assistants

Shawon Sarkar

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:

Chirag Shah, Chair

David Hendry

William Howe

Ryen White

Program Authorized to Offer Degree:
Information Science

University of Washington

Abstract

An Integrated Model of Tasks and Uncertainties for Designing Task-aware Search Assistants

Shawon Sarkar

Chair of the Supervisory Committee:
Chirag Shah
The Information School

Search behaviors are usually motivated by some task that prompts users into the search process. Complex tasks often initiate long, evolving, interactive search processes with shifting goals and cognitive focus at different search stages. Users' search strategies are influenced by their search motivation, encountered problems, and cognitive focus or state of knowledge at these search stages. However, existing search systems are primarily designed to optimize one request at a time, ignoring the underlying overarching task, shifting task phases and sub-tasks with users' cognitive focus, or even the holistic nature of a task-based search session. Although a set of descriptive and theoretical models of the search process can be found in the literature that characterizes tasks, there is a gap in research focused on exploiting dynamic task characteristics in search personalization processes. More importantly, there is a lack of support for users to complete their tasks in an adaptive, dynamic way. To address this issue, this dissertation adopts a multi-disciplinary, human-centered approach and applies a mixed-methods design-based approach to meet three broad objectives: *first*, develop a conceptual framework for understanding how different types of tasks trigger specific information needs

that can lead to different methods and strategies for seeking different forms of information and information sources and, in the due process, identify any barriers they perceive and potential help they choose to overcome those limitations; *second*, apply new computational models to construct unified task representations using underlying search behavioral signals that can be transferable and functional to any task circumstances; and *Third* make existing search and retrieval systems more responsible and efficient to meet the changing state of users' cognitive focus during the search process by using knowledge gained about users' tasks and problems. Specifically, this dissertation aims to develop a task-information need-strategy-problem-based task representation that can be leveraged in search and retrieval models to provide task-based supports in different information formats, thus empowering users to make informed decisions about different aspects of their lives by providing information more relevant to their current task state. The result of this study is a step towards developing task-aware intelligent systems capable of supporting users at each stage of their complex task-completion process.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Number Page

ACKNOWLEDGMENTS

I respectfully acknowledge that the land on which the Information School stands today is the traditional home of the Coast Salish people, the traditional home of all tribes and bands within the Duwamish, Suquamish, Tulalip and Muckleshoot nations.
To acknowledge these lands is to recognize its longer history and our place in that history; it is to recognize these lands and waters and their significance for the peoples who lived and continue to live in these regions, whose practices and spiritualties were and are tied to the land and the water, and whose lives continue to enrich and develop in relationship to the land, waters and other inhabitants today.

\- *Land Acknowledgment*

I extend my heartfelt gratitude to my advisor Dr. Chirag Shah for an incredible six-year journey. I am immensely grateful for the invaluable opportunities he provided me to pursue my interests and engage in fascinating projects. I often resisted his ideas only to realize later their importance in my growth as a researcher. His unwavering support was valuable, especially during challenging times in my professional and personal life.

I am thankful to my dissertation committee members—Dr. David Hendry, Dr. Bill Howe, Dr. Ryen White, and Dr. David Ribes—I am grateful for their thoughtful advice and insightful discussions.

I am deeply thankful to the exceptional professors I had the privilege of crossing paths with during my graduate studies at the University of Washington and Rutgers, The State University of New Jersey. I sincerely appreciate Dr. Nick Belkin for his close mentorship, instruction, and unwavering support. His guidance has been instrumental in untangling complex conceptual knots I encountered.

I owe a significant debt of gratitude to Dr. Nina Wacholder, Dr. Mark Aakhus, Dr. Smaranda Muresan, Dr. Elena Musi, and Dr. Debanjan Ghosh for their unwavering support during my early graduate research project on sarcasm detection as a Masters student. I am also thankful to Dr. Vivek Singh, Dr. Marie Radford, Dr. Robert Kapra, Dr. Pia Borland, and Dr. Clarita Left Hand-Begay for the excellent education they provided. Additionally, I thank my professors at the University of Calcutta, India - Prof. Anuradha Mukherjee, Prof. Subrata Halder, and Prof. Shubhasis Bandyopadhyay - for their guidance during my undergraduate studies.

I am sincerely grateful to my colleagues Dr. Jiqun Liu and Dr. Matthew Mitsui for collaborating on Study 2. I also appreciate Angela Xu, Antony Xu, Sinan Zhao, and Maryam Amirizaniani at UW for their fantastic data processing and annotation work. I wish them all a productive career ahead.

My past and current colleagues have enriched my experience at the InfoSeeking Lab and Data Lab. Dr. Yiwei Wang, Dr. Manasa Rath, Dr. Soumik Mandal, Dr. Souvick Ghosh, Dr. Jonathan Pulliza, Dr. Dongho Choi, and Dr. Matthew Mitsui never failed to inspire me with innovative research ideas. I am grateful for my fruitful interactions with Dr. Ziad Matni, Dr. Chathra Hendahewa, Dr. Long Le, and Dr. Eric Choi. A special mention goes to my colleagues at the CritQuant group - Leo Stewart, Yim Register, and Yubing Tian - and my iSchool research buddies Nicole Kuhn, Josephine Hoy, and Tessa Campbell.

Acknowledgment is due to Victor Aque, Mitchel Ayer, Cheryl Kalent, Kate Kerschbaum, Carmen Parisi, Wendie Phillips, and Dora Tkach from Student Services for their assistance in managing bureaucratic processes and ensuring timely paychecks.

I express my deepest gratitude to my Maryland, New Jersey, Washington, and Kolkata friends and family. The friendship and support I receive from Soumyadip Sengupta, Siddharth Pal, Wrick Sengupta, Arijit Biswas, Pritam Mukherjee, and Ranchu Matthew have been invaluable.

Lastly, I extend my heartfelt appreciation to my parents, Jatindra Nath and Kakali, for their unwavering support throughout my arduous journey in graduate school.

Finally, I sincerely thank my boyfriend and now husband, Dr. Biswadip Dey, who has provided patient support and encouragement throughout this journey. His love and unwavering support have been a tremendous source of strength for me.

I also want to express my sincere appreciation to the Information School at the University of Washington and the School of Communication and Information at Rutgers University, where I had the privilege to work and study.

I am grateful for the financial support received from the National Science Foundation (NSF) for the projects discussed herein. I also acknowledge the financial support the ACM SIGIR and National Science Foundation (NSF) provided for my academic travels.

I apologize to those whom I may have inadvertently omitted, as it is impossible to remember everyone who has contributed to my journey.

Lastly, my deepest and sincerest gratitude goes to Rabindranath Tagore, whose creations resonate with every emotion within me and continue to guide me through the darkest days.

DEDICATION

মা, বাবা এবং বিশ্বদীপ কে

To my mother, father and loving husband

Part I

INTRODUCTION AND BACKGROUND

At the beginning of this thesis, we will give an outline of the structure and explain the contributions of the dissertation. We will present the problem statement and research questions addressed and provide a detailed review of the techniques and components typically used in Information Retrieval (IR) and Interactive IR (IIR), emphasizing the concept of task. Additionally, we will provide background and context for the research problem, including a thorough overview of the theoretical and applied elements that shaped this dissertation.

Chapter 1

INTRODUCTION

A task refers to an activity or a collection of activities aimed at achieving a goal [103]. Tasks encompass various actions, possess distinct limitations, and require different durations for completion [291]. When considering tasks, we typically envision work-related endeavors such as *writing a project report* or *checking work emails*. However, tasks can also encompass everyday routines like *cooking*, or *making restaurant reservations*, as well as leisurely pursuits such as *listening to music*, or *looking for movies to watch*. These activities also involve specific objectives [103]. The nature of people's tasks varies. A task can be complex, requiring a substantial amount of time to complete, or it can be simple and brief [309]. In the realm of search, complex types of search tasks [46] or exploratory search tasks [292] necessitate multiple rounds of interactions with information systems. At the same time, straightforward fact-finding tasks only require a few interactions to obtain information. Furthermore, individuals often perceive tasks in segments and implicitly divide larger tasks into smaller parts [309]. Therefore, a task can be an individual microtask or a component of a larger macrotask [291, 64]. Many extensive and complex tasks, such as *planning a trip*, can be broken down into several smaller, more manageable tasks like *booking flights*, *finding hotels*, or *checking in for flights* [291, 64].

In order to complete a task, it is important to turn its overall objective into specific intentions of what needs to be done. This helps individuals to take the necessary actions [213]. Certain information is frequently required to complete a task or its components, which drives individuals to seek information. Hence, the concept of a task has emerged as a crucial factor influencing behavior related to seeking and searching information, which arises from the need to fulfill specific task goals [43, 100, 277, 295]. Furthermore, tasks provide individuals with objectives to achieve through their information seeking endeavors. Therefore, comprehending the concept of a task is essential in understanding the reasons behind people's information seeking and searching behaviors, the types of information they search for, their preferences for acquiring it, and the contexts in which they use it [44, 252]. Moreover, to grasp tasks fully, it is necessary to comprehend individuals' goals and how they transition from goals to tasks to actions [103].

Given the role of tasks, the Information Retrieval (IR) and Interactive Information Retrieval (IIR) literature extensively discuss and consider various aspects of tasks in information seeking and retrieval, including the topics, types, and other characteristics of tasks [136]. These aspects encompass factors like topic familiarity, which relates to the task itself, as well as user behaviors, specifying functionality, and other characteristics to support users in their tasks [22]. Such aspects significantly influence how individuals perform tasks and search for information. For instance, different task types result in distinct information seeking and searching behaviors [122, 166]. Task characteristics, such as complexity and difficulty, also impact information seeking and searching patterns [41, 47, 100]. Task doers [1] People's information seeking activities can also be influenced by prior knowledge regarding the task topic [3, 35, 107, 140] or familiarity with a specific domain [111, 289, 294].

Earlier research has additionally examined and analyzed the concept of tasks from various angles, encompassing ontological, epistemological, and methodological dimensions. Numerous task typologies or classification frameworks have emerged, focusing on specific aspects of tasks, such as cognitive complexity [138, 301] and task prior determinability [51, 53].

[1] It should be noted that, in this proposal, terms like "user," "information seeker," "information searcher," and "task-doer" are often used interchangeably as the focus of the study is on task-doers who interact with retrieval or filtering systems to fulfill their tasks. The term "user" refers to a system user, such as an information retrieval, search, or recommender system. The term "information seeker" refers to a person, in general, who seeks information without explicitly using a system. A "searcher" is a person who purposefully seeks information on a system. A "task-doer" is one who performs a task.

Another avenue of research presents a holistic and multifaceted framework that defines a task as a composite of several facet values [156, 166]. In the field of IIR research, there is a lot of focus on identifying and representing user tasks. These tasks are defined as overarching goals or problems that guide users' search and exploration of potential solutions. However, a smaller portion of IIR research looks at how users' tasks, as well as their task states or cognitive focus evolve over time, including their information seeking intentions, local search tactics, and encountered problems [173]. When users search for information for complex tasks, their cognitive focus or attention [2] may change as their task progresses. When conducting searches, individuals adjust their strategies based on multiple factors such as search outcomes, interface designs, their own knowledge, task perception, system response expectations, and situational elements.

Understanding the context of a search procedure, such as the task and evolving task states, is pivotal for delivering helpful search support and enhancing conventional retrieval techniques [159]. Hence, examining search tasks and associated user behaviors carries importance in theoretical and practical frameworks, aiding searchers in their endeavors. Consequently, the pursuit of comprehending the task goals and intentions involved in information seeking and searching is not a recent development, antedating advancements in web search, recommendations, and intelligent agents [198]. Nevertheless, understanding users' tasks within the search context remains a fundamental challenge in today's information-driven era. To address this issue, researchers in IR and IIR must devise an analytical, dynamic approach that holds theoretical significance and practical applicability in understanding tasks, particularly complex tasks. In this thesis, we aim to examine the comprehensive depiction of the entire information search procedure and the task behind it.

[2] In this thesis, *cognitive focus* or *task state* is defined as a user's attention towards achieving a specific action or goal during the search process. As searchers acquire more knowledge by submitting queries and reading retrieved results, their cognitive focus changes, and their information needs evolve accordingly [268, 164, 62, 314]. For this reason, a search session is also considered a learning process [99, 173].

1.1 Motivation and Context

People have to do many tasks in their daily lives. In order to complete some macro tasks, as part of those tasks, they often need to collect information by interacting with different information sources, such as interpersonal sources like talking with other people or impersonal sources like search and IR systems. Therefore, individuals interact with search and recommender systems driven by tasks that arise from constantly changing problematic situations in their everyday lives [23, 24]. For many people, commercial web search engines like Bing, Google, Baidu, and Naver serve as the primary sources to fulfill their information requirements [233]. These search systems are designed to act as filters, sifting through large amounts of information and presenting users with the most relevant outcomes, thus making them preferred and dependable information sources for numerous individuals.

Consequently, the suggestions and results provided by web search engines can significantly influence individuals' thoughts, emotions, and actions by shaping their perceptions of accuracy and trustworthiness. Hence, it is crucial to use these systems discerningly and critically evaluate the information they present to us [198]. However, over the past two decades, there has been a substantial transformation in how we search for, understand, and utilize information to accomplish our goals. This transformation is driven by the abundance and accessibility of online information, the widespread presence of social media, and advancements in information and communication technologies. The continuous optimization of search interfaces and algorithms has resulted in improved outcomes tailored to specific queries. The growing adoption of intelligent and conversational agents like Alexa and Siri has opened up new avenues for interacting with information and completing tasks. More recently, large language models such as GPT-3 [220] have revolutionized search and recommender systems by enhancing query comprehension, result ranking, personalization, and content generation. Despite these advancements in search and IR systems, there are still ongoing challenges in addressing the information needs or intents relating to task-solving and goal accomplishment for different users [80, 105]. As a result, the responsibility of managing tasks often falls on the searcher, which can be mentally taxing. There could be various potential causes for this. Often, the main issue lies in the functionalities of the systems, which derive connections between people's cognitive models and intentions from their explicit physical actions [213].

Search and IR systems primarily depend on people's observable manifestation of needs or the verbalization of tasks and their subsequent search actions to retrieve relevant information. Search systems may appear proficient at providing answers, assuming users know what to ask. However, users, particularly novice searchers, often struggle to formulate effective queries due to a lack of knowledge, uncertainty about their needs, or inability to articulate their intentions. While people interact with search systems motivated by tasks that emerge from evolving, continuous problematic situations [23], they often have difficulty articulating their goals, information needs, and intentions behind the tasks while interacting with the systems (1.1). As Donald Norman pointed out, "For many everyday tasks, goals, and intentions are not well specified" [213], sometimes the goals are vaguely stated, and intentions lack specificity. This often stems from a lack of clarity and knowledge about tasks or how a search or retrieval system operates, or even due to the overwhelming volume of information and the resulting cognitive and affective stress [287]. Additionally, people may not be aware of what they do not know while performing a task [252]. All these factors can result in users expressing ill-defined tasks as queries to the systems. So that can create a gap between users' actual task and the task perceived by the system by relying solely on users' expressions of needs through questions or queries; hence that cannot guarantee an infallible approach.

Furthermore, existing search frameworks have primarily focused on handling discrete, transactional, straightforward, navigational, or factual searches individually [288], assuming that addressing individual interactions will suffice to fulfill the overall task of the searcher. However, this approach disregards the fact that users' overall tasks and objectives often drive search behaviors. Several researchers have pointed out that flawed assumptions (e.g., [20,

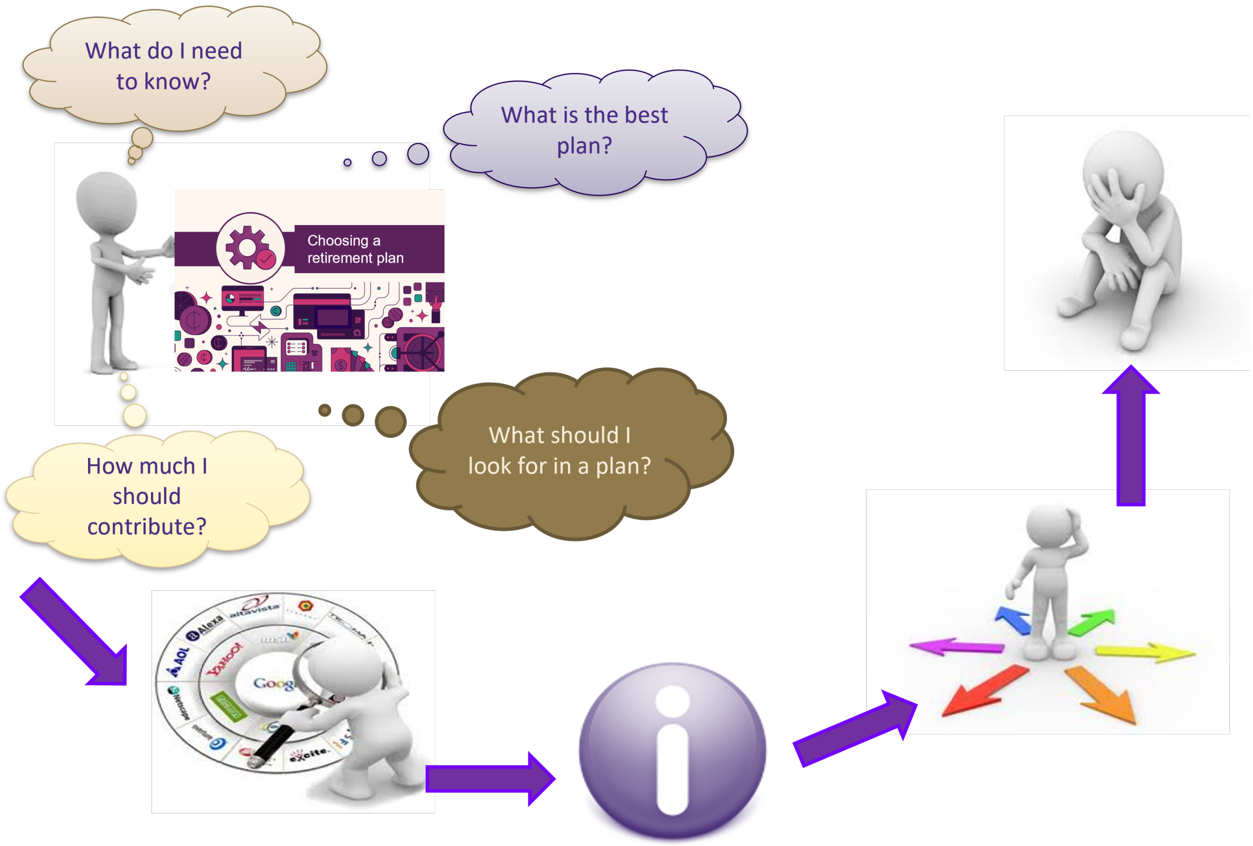

Figure 1.1: Individuals encounter numerous challenges when seeking information to accomplish those tasks, including a deficiency in comprehending the information requirement, unawareness of their knowledge gaps, a vast amount of information and information saturation, unfamiliarity with the functioning of information retrieval systems, and cognitive and affective stress.

106, 198, 171]) that places the cognitive burden of monitoring various task aspects (e.g., exploring the knowledge domain, identifying essential sub-tasks, and formulating queries to accomplish these tasks) on the users themselves, particularly for complex tasks requiring assistance for longer and often across multiple search sessions. Current search systems lack an understanding of the underlying task and often fail to consider the type of task a person is working on when optimizing results or suggesting what actions a person should take [106], resulting in search results that may not directly align with users' local and overarching task goals. This limitation becomes evident when attempting complex, exploratory tasks that involve problem-solving and decision-making. Such tasks demand tools and systems beyond single-query or single-turn interactions, providing support at the task level throughout the search sessions [143, 166, 164]. Therefore, as search tasks become complex, the system's level of support decreases over time. The complexity of the search process poses significant challenges for information seekers with limited information literacy or novice searchers who lack knowledge about the task they are trying to complete. Additionally, systems offer limited options for exploring and interacting (in terms of related query and document suggestions) with information and provide little assistance in consolidating and integrating information across multiple search sessions, approaches, and modes [198, 207]. For example, queries like *What is the distance between Earth and the Moon in kilometers?* can yield an instant answer at the top of the search engine results page (SERP). Even brief yet comprehensive details about specific individuals, groups, or entities, such as for query *BTS*, are often presented in condensed cards that display information like age, birthplace, family connections, associated songs, and recent news [34]. However, when it comes to exploratory intellectual tasks like determining *whether pursuing a degree in computer science or information science is worthwhile*, multiple rounds of browsing and reading results are often required to find satisfactory answers.

Although efforts have been made from the systems' side in query recommendations and auto-suggestions to assist users [153, 273], there are inherent challenges in operationalizing complex search needs. Complex tasks, for example, determining the *worthiness of pursuing a degree in Information Science or Computer Science*, cannot be addressed with a single query. Various sub-tasks and related goals need to be explored, including clarifying concepts, understanding philosophical differences, and investigating relevant topics and courses in each program. As pointed out above, the challenge lies in the systems' ability to derive the relationships between users' tasks, intentions, actions, and articulation into queries or physical search actions [213]. Search and IR systems optimize users' queries, but searchers' tasks - not their queries - ultimately drive their searching behavior. These systems rely on people articulating tasks into words or their search actions to provide useful information.

In recent years, advancements in retrieval systems and conversational agents leveraging large language models (LLMs) have shown promise in supporting longer search sessions and complex tasks. LLMs, such as PaLM 2 [3] powered Google Bard and Open AI's GPT-4 [4] powered BingChat, can generate contextually relevant responses, leading to extended interactions and deeper exploration of information. These models better understand user intent and context across multiple search sessions, providing detailed and informative responses that accommodate follow-up questions or requests for clarification. They can also remember and reference previous interactions, enabling continuity and coherence throughout a more extended search session.

Nevertheless, challenges persist despite the progress made by generative AI and LLM-based search systems. As Shah and Bender pointed out [253], these systems can still only support a few information seeking strategies [26]. These systems may occasionally produce incorrect or biased information and struggle with understanding complex or nuanced task objectives. They often consolidate search results into a single definitive response, limiting users' ability to explore and browse through the results. LLMs excel at generating prescriptive responses based on intricate probability distribution models, which can undermine users' freedom to consider diverse options and exercise their own judgment [94, 130]. This lack of judgment can lead to biased results, reinforcing societal biases. Additionally, LLMs may prioritize specific sources and overlook less prominent or diverse perspectives, limiting their ability to provide a comprehensive overview of relevant information. LLMs-based search

---

[3] https://ai.google/discover/palm2/

[4] https://openai.com/gpt-4

systems have yet to grasp the nuanced context of search processes fully [202, 253, 30], which necessitates human reasoning beyond straightforward execution. These systems struggle to infer accurately and cluster users' search intents and task goals, resulting in a less personalized search experience. Consequently, they may not effectively assist users in task completion, especially when faced with ambiguous queries, which can lead to inaccurate or irrelevant responses. Addressing these limitations requires ongoing research and development efforts. Understanding users' overall tasks and tracking the evolving states of tasks and problems throughout search sessions can enhance the capabilities of LLM-based retrieval systems and address their limitations.

1.1.1 An Example of Complex Search Session

Figure 1.2 and Figure 1.3 present example complex search task sessions taken from our second user study (Chapter 5) conducted on a popular commercial search engine and the other with an intelligent search agent, respectively. To complete this complex search task, searchers must gather information on various aspects of the task. These examples highlight the limitations of current search and IR systems in understanding and supporting complex search tasks. In Figure 1.2, the searcher performed multiple queries and information evaluation rounds to achieve different goals at different query iterations. As the search progressed, the searcher continuously interacted with the system and search results; they gained more knowledge about their tasks and encountered different challenges, and their cognitive focus (states) kept evolving (e.g., from exploration state to exploitation state) like a learning process. As a result, their information needs constantly shifted in every query segment.

Figure 1.3 shows the one definite answer provided by the LLM-based systems.

In both search sessions, the searcher faced challenges completing their tasks and exhibited signs of lacking domain and topic knowledge while formulating queries [242]. However, the systems failed to understand the struggles and provide appropriate support. They relied

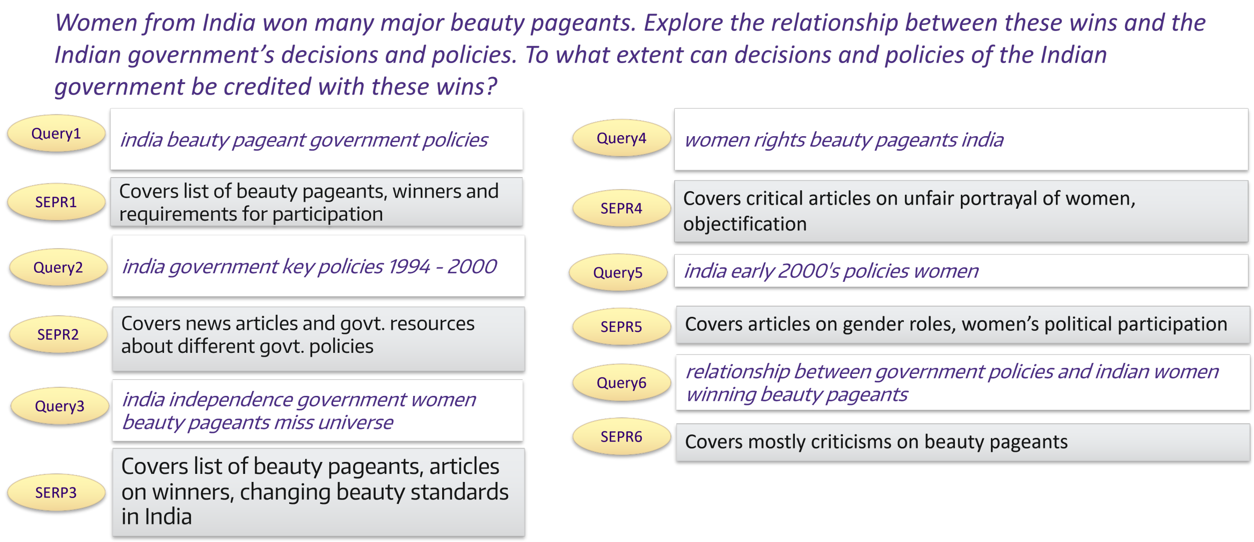

Figure 1.2: Here is a summary of a complex search task session performed by a searcher on a search engine with queries submitted by the searcher and a summary of SERPs chronologically returned by the system. The paragraph at the top summarizes the overall goal of the task.

heavily on the explicit expressions of search intents, even when expressed vaguely, leading to the retrieval of less helpful information. The systems did not facilitate the improvement of query formulations or the understanding of the overall task goals. As a result, the cognitive burden of keeping track of the evolving task fell on the searcher, creating a gap between their cognitive focus and the perceived mental state by the systems.

In order to help users in situations where they need assistance, there is a requirement for intelligent support that can identify and carry out specific tasks. This system should be able to recognize the challenges users face during different stages of their tasks and provide appropriate assistance to help them with their search activities or any other activities that involve searching [242, 243]. So it is critical first to understand those activities and how the search fits in. The effectiveness of a search system should be measured based on its ability to resolve users' information problems and help them achieve their goals [198]. To do so, we

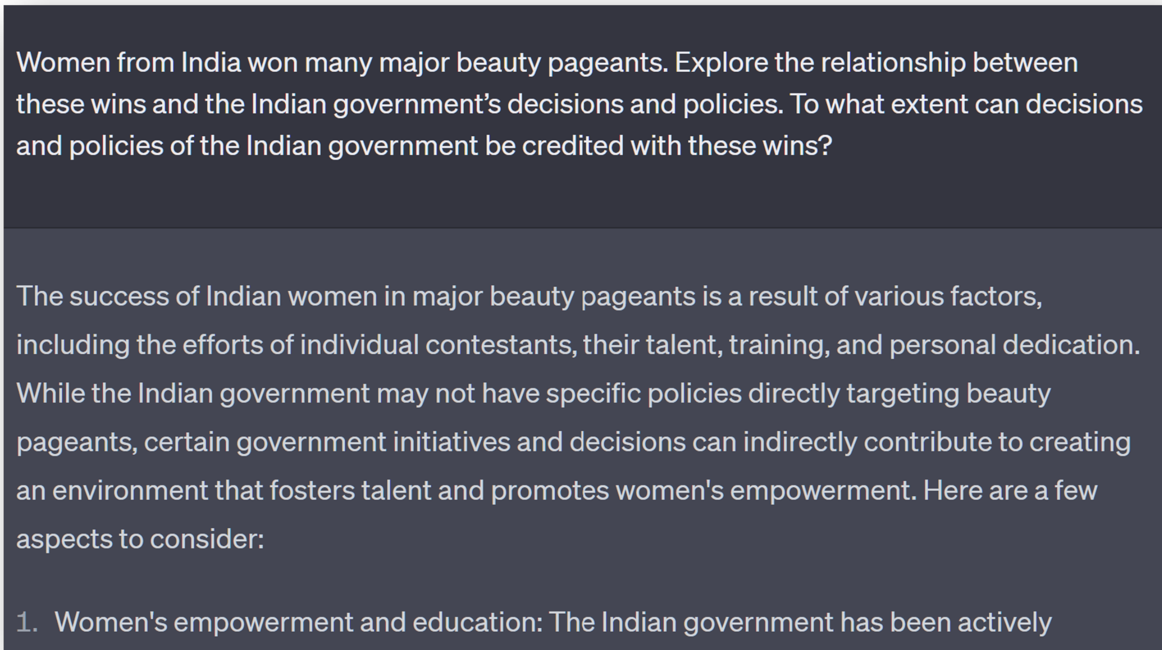

Figure 1.3: Here is an example of a complex search task session conducted by a searcher on a conversational search chatbot. The chatbot provided a single synthesized answer.

need to address the following components [87]:

- capturing and representing search tasks at various levels;
- capturing and representing latent task states that emerge during user interactions to support their changing intents;
- incorporating these task representations into search or recommendation applications;
- deciding on appropriate intervention types and timing that do not diminish the agency of searchers;
- identifying helpful information types and modalities to provide intervention support.

In this thesis, we address the first three components mentioned above.

1.2 Problem Statement and High-level Research Questions

With the above discussion in mind, this dissertation research is guided by the insights from previous studies on tasks and adopts a multi-disciplinary and human-centered approach to achieve three main objectives, as follows:

Objective 1: Developing a Conceptual Model The first objective is to develop a conceptual model that comprehends how different types of tasks elicit specific information needs, leading to diverse methods, strategies, and information sources for seeking information. The aim is to understand the users' tasks, the motivations behind task needs, their search approaches, and the challenges they encounter during each step of their search sessions [259, 243]. In other words, we aim to construct a task-information need-strategy-problem map, which can be utilized to offer task-based support in various information formats (e.g., suggesting queries, documents, or relevant individuals) to address the encountered problems and guide users towards successful task completion.

Objective 2: Understanding Evolving Task States The second objective is to explore how the knowledge of evolving task states, underlying search needs, and the overall goal can enhance a system's offerings to its users. Addressing the challenges mentioned above necessitates automated methods for the sense-making of tasks and the development of task-based information retrieval systems. These systems should excel in modeling, identifying, and extracting tasks while also supporting users' complex search tasks across various search situations, modes, and modalities. Building upon the insights gained from the first objective, this dissertation aims to develop computational models that can extract the nature of users' tasks from their search behaviors.

Objective 3: Leveraging Task Knowledge in Search Applications The third objective focuses on utilizing the acquired task knowledge in different search applications. Specifically, the emphasis is on learning task-based responsible design for scalable and efficient information systems. These systems aim to align with users' task goals and deliver relevant information in various formats, such as queries, documents, and people. By understanding and leveraging task information, this research aims to address the challenges users face during the search process and provide effective support to overcome them.

1.2.1 High-level Research Questions

After analyzing the problem space and outlining our objectives, we have crafted four overarching questions to explore in this thesis.

RQ1. To what extent do various task types trigger distinct information needs, prompting diverse approaches and strategies in the search for different types of information and sources?

RQ2. What problem(s) do searchers encounter while performing a specific task, and the type(s) of help do they prefer to get from the system?

RQ3. What approaches can be employed to effectively represent and express tasks in a manner that is search context agnostic?

RQ4. In what ways can we integrate context-agnostic task representations into current search and retrieval approaches to enhance their efficacy?

Since information seekers do not make decisions regarding their information-seeking process in isolation but rather as part of a holistic process driven by tasks, our first research question (RQ1) examines the relationships between core aspects of task-based information seeking behavior and the various dimensions of these aspects. These dimensions include "what" (nature and form of information), "why" (information need), "how," and "where" (information source). By understanding these relationships, we can gain insights into the information seeking process.

The second research question (RQ2) focuses on investigating the challenges or issues individuals encounter while searching for information and the types of solutions they prefer to receive from information systems. This knowledge will enable systems to identify users who are facing difficulties during their search processes and offer appropriate interventions based on their preferred assistance. Thus, these first two research questions, aligned with our first objective, seek to comprehend the dynamic nature of task states and the evolving problems users encounter during the search process. The answers to these questions contribute to advancing our understanding of representing and explaining the dynamic aspects of complex search tasks.

Moving forward, the third research question (RQ3) explores the relationship between observable search behaviors and implicit task information in search interactions, intending to extract tasks from these behaviors. By understanding this connection, we can enhance the practical applications of our research.

For practical applications, the fourth research question (RQ4) delves into the possibility of leveraging knowledge about users' tasks and challenges to improve search efficiency and provide adaptive support for users engaged in complex search tasks.

Considering the benefits of a context-agnostic task representation, such as flexibility, generalization, scalability, transferability, and ease of implementation, it becomes apparent that this approach is valuable for enhancing the effectiveness of search and retrieval systems across diverse tasks and contexts.

Taking the aforementioned points into consideration, the primary argument put forth in this thesis is that by taking into account the dynamic nature of users' task states and the varied search challenges they encounter at different stages of the search process, we can construct more reliable and authentic models that accurately represent the intricacies of the search process. This knowledge is a valuable means for gaining deeper insights into user tasks and the complicated interactions between users and the search systems they utilize. The insights derived from this research can then be utilized by researchers to develop more intuitive and realistic search/IR models and evaluation metrics that facilitate the assessment of task-based retrieval systems and their users, thereby advancing the field of interactive information retrieval.

1.3 Thesis Contribution

In this thesis, several significant contributions are presented. We categorize these contributions into primary contributions from conceptual and theoretical perspectives, algorithmic and methodological advancements, empirical insights, and broader implications.

1.3.1 Conceptual and Theoretical Contributions

Our first contribution involves the development of a new task model. Building upon existing task models (e.g., [173]), we propose an updated search process model that emphasizes the evolving task states or users' cognitive focus during their search journey. This addresses the research question RQ3 in Chapter 6. The conceptual task model, outlined in Chapter 6, encompasses various aspects and decision points that searchers encounter throughout their search process (Chapter 4), establishing a flow of interaction based on established models. This enhancement allows for a more comprehensive understanding of the complexities involved in the search process. As a conceptual model, it offers the flexibility to be operationalized in different ways, enabling its practical application.

Furthermore, numerous challenges are identified in the literature that users face while seeking and searching for information. Our second contribution, discussed in Chapter 5, is a task-problem-help mapping. This framework focuses on inferring potential problems users may encounter during their search and recommends suitable assistance to mitigate those problems across different task scenarios. Given the limited empirical work on understanding the barriers faced by searchers in various task scenarios, this framework provides system designers with valuable insights to better model user interactions and identify their struggles, ultimately facilitating the provision of appropriate support interventions.

1.3.2 Algorithmic and Methodological Contributions

The first contribution involves operationalizing the conceptualized unified task representation model that utilizes latent search session structures. This model focuses on complex tasks and the evolving task states, addressing research questions RQ3 (Chapter 6, Chapter 7) and RQ4 (Chapters 9 and 10). It complements the first conceptual contribution and provides a global context-agnostic approach to task representation. This approach supports diverse user interactions and task management. Additionally, the ability to model users' tasks from their interactions opens up new possibilities for solving various problems and enhancing user engagement and satisfaction. The dissertation incorporates graph neural networks, particularly metapath-driven heterogeneous neural networks, into the task research to accomplish this.

The second contribution is the development of a general and straightforward workflow for simulating synthetic search session behaviors on a large scale (Chapter 8). This workflow may prove valuable for researchers who may not have the necessary resources to directly collect user data or undergo a time-consuming data annotation process. The ability to simulate search session behaviors at scale enables researchers to explore and analyze user interactions more comprehensively, aiding in developing and evaluating interactive information retrieval systems.

1.3.3 Empirical Contributions

Our dissertation focuses on using our task representation model to improve search and recommendation problems. We conducted two experiments to explore this implementation. We found that understanding the relationship between task characteristics and search behaviors is crucial. By studying this relationship, we discovered that knowing users' search behaviors and specific task characteristics can enhance retrieval and recommendation quality. Additionally, task characteristics and user behaviors affect decision-making behaviors and can potentially influence each other. This emphasizes the need for a comprehensive approach when studying tasks in the field of IR and IIR.

This dissertation contributes significantly to the field of IIR by examining search interactions during complex search tasks and linking task-based information seeking theories to empirical practice. The implementation of task models in search and retrieval models emphasizes the importance of considering tasks as contextual features in IR research. By considering users' changing cognitive focuses, the dissertation creates, implements, and evaluates cognitive focus-based models of complex search tasks that describe and explain the dynamic nature of these tasks. Furthermore, it investigates the relationship between users' task states and search tactics, paving the way for the development of task-aware adaptive search supports. Precise task representation based on user-system interactions aids in understanding users' requirements, allowing systems to provide superior recommendations and personalized experiences. This approach addresses the challenges users face while searching and goes beyond ad-hoc retrieval, providing support to achieve overall goals or tasks associated with search sessions.

1.3.4 Broader Contributions

The results of this study can have implications for both work and everyday life situations. By using advanced LLM-based systems that include task-based information retrieval and recommendations, individuals can have improved access to information, overcome barriers they face during information seeking, and make more informed decisions that contribute to their overall well-being.

Incorporating a user's task information into LLM-based search systems can provide enhanced support for task completion that goes beyond traditional transactional search systems. This integration allows the system to align its responses and recommendations with the user's specific task goals, resulting in a more personalized search experience. Additionally, by considering the user's task information, systems can generate recommendations and responses that are more contextually relevant and aligned with the user's intent.

Task-based approach not only improves the effectiveness of the search system in facilitating task support and completion but also enhances interpretability and explainability. Users are provided with a clear understanding of why specific recommendations are being made or certain search results are being presented, improving their trust in the system.

By including task information in the system, a better understanding of the user's information needs and the context in which they arise can be achieved. This results in the system being able to offer more helpful and valuable support throughout the entire task lifecycle rather than just fundamental transactional interactions. The system can offer guidance, relevant resources, and assistance at every stage of the task, considering the user's changing requirements and preferences. This approach increases user satisfaction and a more productive and effective search experience.

1.4 Thesis Outline

This section provides a brief summary of the parts and chapters of the thesis.

Part I introduces the research problem, motivation, background and contributions of the thesis.

> **Chapter 1**, within the broader context of this dissertation, serves as an introduction to the main focus of the thesis. It provides an overview of the research questions, contributions, and the overall structure of the work.
>
> In **Chapter 2**, the background and prior work in the fields of IR and IIR are

discussed. The chapter begins by explaining the fundamental concepts of IR and IIR, specifically emphasizing how models and measures consider tasks in information seeking. It then reviews existing literature that examines users' task behaviors, highlighting established IR concepts, search models, and methods in IIR. Various user behaviors defined in the literature are outlined, followed by an exploration of evaluation measures used in task-based IIR. The chapter also explores the link between user characteristics and task behaviors. It discusses recent research progress on task-based information seeking, users' search problems and needs, and existing theories based on empirical evidence.

Part II, presents our empirical contributions. In this part, we present the user studies that were undertaken, as well as the experiments we conducted that allow us to address research questions.

Chapter 3 presents an overview of the workflow of the thesis.

In **Chapter 4**, the characterization of tasks as a multidimensional construct is explored, utilizing a user study and a few experiments to understand the relationships among users' interactions in different task contexts, addressing **RQ1**.

In **Chapter 5**, a user study is reported, investigating users' task-level behaviors, problems, and preferred assistance, addressing **RQ2**.

The beginning of the **Chapter 6** introduces the conceptual problem-help-task model for task states, emphasizing task characteristics and user behavioral signals. The assumptions of the model are discussed. The latter half of the chapter and **Chapter 7** studies user search behavior and presents algorithms for extracting search tasks from logged interaction data, addressing **RQ3**.

Before delving into **RQ4**, we focus on developing a workflow to simulate synthetic search sessions with task information labels for large computational models to overcome the limitations of small-scale training datasets. In **Chapter 8**, we report on our approach to create the simulated search sessions.

The final section of this part focuses on how to use the task information extracted in various applications. We conducted two experimental scenarios - search and recommendations - to test our task representation model and determine if it enhances the performance of search and recommender systems. The last two chapters of this part describe how we implemented the task representation model in existing search and retrieval systems and several simulated experiments that answer research question **RQ4**. In **Chapter 9**, we present a task-based re-ranking model.

In **Chapter 10** discusses a novel task-based recommendation model which leverages task context to learn top-k recommendations.

Finally, **Part III** of this thesis consists of the concluding chapter.

Chapter 11 summarizes the work undertaken, implications of the main findings, potential assumptions, and limitations. It also discusses avenues for future research.

Chapter Summary

- Tasks encompass a wide range of activities aimed at achieving specific goals, including work-related tasks, everyday routines, and leisure pursuits.

- Tasks play a crucial role in information seeking and searching behaviors, as individuals seek information to fulfill specific task goals.

- The characteristics of tasks, such as complexity and difficulty, influence information seeking and searching patterns.
- Understanding the context of a search procedure, including tasks and evolving task states, is crucial for effective search personalizations.
- Representing and operationalizing task knowledge within the search systems remains a fundamental challenge in IR and IIR research.
- Advancements in technology have transformed how we search for and utilize information. However, there are ongoing challenges in meeting the information needs related to task-solving and goal accomplishment.
- Current search systems often rely on users' explicit expressions of needs and struggle with handling complex tasks. On the other hand, users often need help articulating their goals and intentions while interacting with search systems.
- The thesis aims to advance the understanding of task-based retrieval systems and complex user-system interactions.
- The thesis aims to integrate user task information into search and recommendation systems that yield several benefits, including improved support for overall task completion, enhanced interpretability and explainability of recommendations, and a more comprehensive and tailored search experience.
- By leveraging task information, these systems can go beyond transactional interactions and offer users meaningful assistance, ultimately facilitating their task goals and enhancing their trust in the system.

Chapter 2
BACKGROUND AND LITERATURE REVIEW

This dissertation primarily focuses on the intersection of human information behavior (HIB), interactive information retrieval (IIR), information retrieval (IR), and human-computer interaction (HCI). Figure 2.1 visually represents where the research is positioned within these interconnected areas. Additionally, this chapter presents the theoretical, conceptual, and methodological background, along with related work, that informs the research conducted in this dissertation.

Research on people's information behaviors and their interactions with IR systems can be categorized into two distinct lines. First, information retrieval (IR) research draws from the system-oriented Cranfield paradigm [71], focusing exclusively on the system itself. Second, Interactive Information Retrieval (IIR) research primarily focuses on users' behaviors, tasks, and information needs rather than the requirements of IR systems [71]. IIR research bridges the gap between system-oriented and user-oriented approaches, ensuring that IR systems are accessible, usable, and effective for users [234].

2.1 Information Retrieval

The field of Information Retrieval (IR) research has undergone significant evolution since its establishment in the 1950s, aligning with the development of computers, the Internet, and intelligent communication technologies [185]. This progress has facilitated the study and experimentation of IR, leading to the emergence of various retrieval models grounded in diverse theories and accompanied by methods for evaluating their effectiveness. The primary objectives of IR studies revolve around indexing and representing document retrieval and

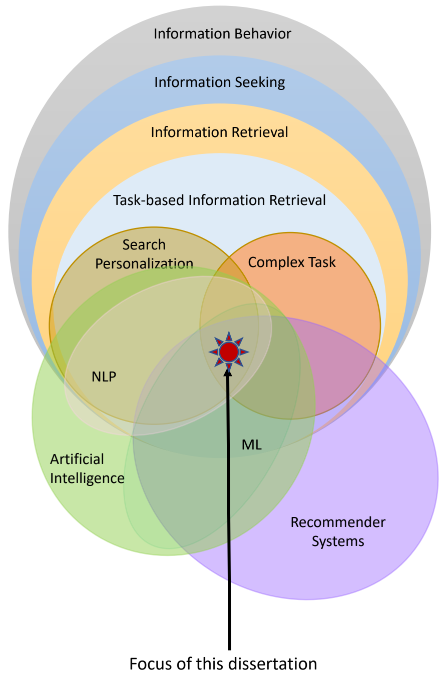

Figure 2.1: Situated Literature Review

retrieving them efficiently [185].

Information retrieval, as defined by Manning [185], is the process of locating unstructured material, typically textual documents, that fulfill an information need (topical similarity) from extensive collections stored on computers. IR research aims to reduce information overload and search times [237]. Initially, the IR process relied on manual library-based approaches. However, with increasing computational power and storage capacity, it shifted towards more automated methods [198]. By the 1970s, efficient retrieval techniques were demonstrated on small text corpora like the Cranfield collection [71], which comprised several thousand documents. The field of IR experienced rapid advancements driven by the need to search for information within ever-expanding volumes of data generated and stored. The advent of the Internet and the accessibility of digital technologies further accelerated this progress. The Cranfield experiments [71] played a crucial role in the development of IR. They established a standardized experimental paradigm that employed the same set of documents and information needs across languages and utilized standard IR measures such as precision and recall to evaluate system effectiveness. However, these experiments made simplifying assumptions; for example, the notion of topical similarity was approximated to determine relevance, all relevant documents were considered equally desirable, and the relevance of one document was assumed to be independent of others [281]. Additionally, Cranfield assumed a static information need, disregarding any changes in the searcher's search process [281].

Recently, IR research has been actively exploring and integrating advanced language models and graph models to enhance retrieval performance. Large language models such as BERT [86], GPT [1], PALM2 [2], and Transformer-based architectures have shown promising results in natural language processing jobs and inferring users' search intents (e.g., [81, 186]). Researchers are investigating how to leverage these models to improve document representation and query understanding in IR systems. Moreover, graph models like graph neural networks and knowledge graphs are being employed to capture rich semantic relationships between documents, entities, and concepts (e.g., [299, 91]). These endeavors could revolutionize the field of IR and enable more effective and personalized information retrieval experiences for users.

2.2 Interactive Information Retrieval

Interaction Information Retrieval (IIR) research, as discussed by Kelly [137], explores the interactions between users, search systems, and information. It encompasses the analysis of users' actions, thoughts, and emotions when using search systems. Two main areas of study

[1] https://openai.com/gpt-4

[2] https://ai.google/discover/palm2/

influence the field of IIR. Firstly, there are investigations into how people search for and seek information. Secondly, there are efforts to develop innovative methods for interacting with electronic resources [234].

The primary objective of IIR studies is to assess the effectiveness of a given search system in assisting users in finding relevant documents [137]. These studies often involve comparing the performance of different search systems or examining how contextual and situational factors impact users' cognitive processes and search behaviors. Researchers typically assign tasks to study participants to facilitate comparisons and systematic evaluations, enabling a comprehensive analysis of their interactions.

IIR research encompasses aspects from both user-sided and system-sided perspectives. For instance, researchers may present the outcomes of a user study that explores a specific aspect of a searcher's behavior while also providing details of a system evaluation. Kelly [137] suggests that IIR has its roots in various disciplines, including traditional information retrieval, library and information sciences, psychology, and human-computer interaction. While often viewed as a branch of IR and/or HCI, there are arguments supporting the consideration of IIR as a distinct area of research [234]. To better define the position of IIR within the system-sided and user-sided space, Kelly [137] presents a spectrum of work that bridges IR and IIR. This spectrum consists of eight different categories of study, ranging from solely system-sided investigations, such as TREC-style studies, to those that focus more on the user's perspective. The categories also include log analysis, which considers a searcher's behaviors when interacting with a retrieval system.

2.3 Information Seeking and Searching

The characteristics of the four primary information behaviors, namely *information seeking*, *information searching*, *information acquisition*, and *information use*, are complex and can be interpreted differently [251].

Information seeking involves the process of an individual resolving their information need [251]. It can be considered a middle-level category situated between macro-level constructs like human information behavior and micro-level categories such as information search or information retrieval [251]. Initially defined in the 1970s as specific actions aimed at satisfying information needs [251], information seeking gained attention through Wilson's article on user studies and information needs in 1981 [297]. Wilson proposed that information seeking behavior arises when an individual perceives a need and interacts with information sources or services, resulting in success or failure to find relevant information [297].

Information searching behavior, a subset of information seeking and a micro-level behavior, refers to purposeful actions when interacting with information searching systems, including information retrieval systems [16, 18]. It is one of the modes of information seeking and generally involves active and directed browsing, monitoring, and reading of information sources to fulfill specific information needs [16, 18].

Information acquisition is another mode of information seeking and involves obtaining information [295]. On the other hand, information use encompasses the physical and mental acts of incorporating the information found and how users perceive the helpfulness of that information in addressing problematic situations [295].

2.3.1 Components of Information Seeking and Searching

According to Kuhlthau [148], information seeking involves fitting information with what one already knows and extending that knowledge to create new perspectives. Past studies have identified the following fundamental factors involved in individuals' information seeking: information/content/material, information needs, sources of information, and methods to access information sources. This section covers different dimensions for each of these constructs are identified. From the theoretical perspectives, terms such as information seeking, tasks, and information needs are contentious and difficult to agree on among researchers;

therefore, we define the concepts for this study and overall dissertation research purposes.

Information

The term “information” is typically used to represent various overlapping and contradictory concepts, and the meaning is varied based on different philosophical traditions. Two dimensions of information construct have been presented in the existing literature:

- the representation/manifestation or expression of information such as
  - communicated knowledge perceived by the cognitive state of mind or change in existing knowledge structure [83, 38, 21, 195],
  - sensory stimuli [231, 182],
  - the process of gaining knowledge [218, 176, 39, 195],
  - the process of communication [261],
  - social construct [78, 231, 54],
  - and objects conveying information [39, 231, 195]
- the interpretation (form or nature of information) of above representation (e.g., [39, 255, 190]) such as
  - opinion,
  - advice,
  - fact,
  - social or emotional support

Each of these dimensions of the information construct can play a more significant role than the other depending on the information seeking context and based on other critical attributes of a seeking process such as the source of information, the seeker, and the problematic situation that triggers the information need. Therefore, one can interpret "information" as something meaningful that can add or change existing human knowledge structure, a form of a tangible object, or a form of a thinking process based on what seekers define or seek information in a particular context or situation.

Information Needs and Motivations

The concept of information needs holds significant importance in the field of IIR. Information needs exhibit specific characteristics based on their utilization. They can be seen as a means to obtain answers, reduce uncertainty, or address existing gaps in knowledge [59]. As users progress through the information seeking or searching process, their needs evolve, and their perceptions of information change accordingly, reflecting their increasing understanding and knowledge of the problem at hand [238, 146]. Taylor [267] proposed a cognitive perspective on visualizing information needs, outlining the stages involved in the mental process. He described four stages in the construction of information need: the visceral need, which is the unexpressed need experienced by a user; the conscious but inexplicable need; the formalized need that can be expressed in natural language; and finally, the realized need, which is presented as a question or query to a source [267]. Building upon Taylor's theory, Belkin [19] introduced the concept of the Anomalous State of Knowledge (ASK) as a starting point for the information search process. Belkin defined information need as an uncertain and unspecified problem arising from a gap between the user's knowledge about the problem and what they need to know to solve it [19].

From a different perspective, Wilson [296] conceptualized information need as a secondary need dependent on three primary interrelated human needs: physiological, cognitive, and affective. These primary needs arise from the social environment and individuals' roles in their social and work contexts. Wilson emphasized that information needs emerge in sociological environments as problematic situations arise and information is sought and utilized. The cognitive state of individuals, socio-cultural settings, and situational factors (e.g., tasks) contribute to the variability in information needs [296]. To define the concept of information need, it is crucial to consider how users perceive and articulate their needs, interpret the task situations triggering their information needs, and search for and use information while considering the social and physical environment. Katz et al. [131] categorized a number of basic human needs into five groups: (i) needs for strengthening information, knowledge and understanding (cognitive factors), (ii) needs for strengthening aesthetics, pleasures and emotional experience (personality factors), (iii) needs for strengthening credibility, confidence, stability (judgment of information values and relevance), (iv) needs for strengthening contact with family, friends and the world (situational and social and environmental factors), (v) needs for escape or tension release (affective factors).

Moreover, affective aspects, such as motivations and expectations, can also undergo transformation and impact the evaluation of information and its perceived need [146]. Motivation consists of the "inner states" that guides or prompt behavior [183], meaning that individuals are compelled to act in specific ways in order to satisfy their information needs. Motivations can be objective, leading users to seek factual information to make informed decisions or solve problems [9] or subjective, perceiving information needs as socio-cognitive requirements when individuals recognize problematic situations or knowledge gaps and attempt to fill them by seeking relevant information [59]. Zhang [312] categorized motivations into three broad factors: cognitive, social, and emotional, taking into account the users' perception and translation of their information needs. Similarly, Oh [214] identified ten motivational factors: enjoyment, efficacy, learning, personal gain, altruism, community interest, social engagement, empathy, reputation, and reciprocity. Based on Katz et al. [131], Choi and Shah [65] created a comprehensive list of motivational factors and sub-variables to investigate different motivations that drive people to seek information while there is a need to address a task:

- Cognitive needs
  - Finding relevant information in immediate surroundings, society and the world
  - Seeking advice or opinions for making decisions
  - Learning; self-education through acquiring information
  - Gaining a sense of security through Knowledge
- Affective needs
  - Looking for social and emotional support for personal issues
  - Looking for social and emotional support for someone (e.g., family, friends, etc.)
  - Looking for attainment on personal thoughts or Ideas
- Personal integrative needs
  - Finding support for one's own values
  - Gaining insight into one's own life
  - Experiencing empathy with problems of others
- Social integrative needs
  - Identifying with others and gaining a sense of belonging
  - Finding a basis for conversation and social interaction
  - Having a substitute for real-life companionship
  - Feeling connected with other people

- Tension free needs
  - Having fun asking a question on Yahoo! Answers
  - Filling time
  - Emotional release

In this dissertation, we utilize the categories of motivations behind information provided by Choi and Shah [65] and Katz et al. [131] as their frameworks are the most comprehensive ones.

Information Source

Individuals utilize various sources available to them while fulfilling their information needs, such as friends and families, books, and websites. An information source is a repository that stores and provides knowledge or information [68, 307]. Sources of information can be impersonal or non-human such as online sources, physical documents, various information retrieval systems, or interpersonal or human-related such as friends and colleagues on various online or offline channels [1]. Several studies on information source selections and preferences have also found that different kinds of information sources are preferred in different situations, in different everyday life and work-related task contexts (e.g., [247, 149, 235, 236, 47]). Moreover, there are various conditions, including seekers' previous experience with a source, accessibility of the source, and the format and content of the information available [158], by which information seekers value and select information sources. Anderson [4] found that people perceived information sources that are easier to use as more accessible and choose them frequently regardless of the quality of the information they expect to obtain. On the other hand, Ashford [7] found out that the source quality is more important than accessibility while thinking in a cost-benefit framework. Besides, time plays an influential role in selecting the type of source [287]. Fisher [93] identified several preference criteria for information sources in their study, such as trustworthiness, contact, access or convenience, inexpensiveness, and ease of use.

Methods to Access Information Source

Although information sources and the methods or channels to access the sources are often synonymous in many existing studies [47, 60], these two concepts are distinct [1]. From existing research, the study identifies two dimensions of the method of accessing information sources and adds another dimension of mediation involved in accessing the sources:

- Physical-electronic dimension - the use of a physical or electronic medium for information transfer [1]
- Synchronous-asynchronous dimension - the synchronicity of communication or connection [1]
- Mediation-No mediation - involvement of any mediated entity

According to Xu [307], channels are the modes of communication through which content is delivered from an information source to seekers, such as face-to-face, phone, or e-mail. Various cues [302] shapes individuals' perceptions of information sources.

2.3.2 Problems and barriers in search interactions

Previous studies have shown that individuals frequently face difficulties when looking for information (e.g., [92, 248, 249, 287, 242]) which cause uncertainties during an information seeking searching process [67]. These include internal barriers (e.g., lack of knowledge, unable to articulate and express the need) [27], external barriers (e.g., time constraints, institutional restrictions) [245], and interpersonal barriers (e.g., lack of help from other people) [265], and other barriers as well [67]. In their study, [67] identified twenty-one problems or barriers people might face during information seeking processes, such as information overload or irrelevant search results. When people encounter problems in the search process, they desire help or assistance either from an IR system or a human to solve problems [117]. Xie [305] investigated help-seeking behaviors or problematic situations that arise in searching digital libraries and identified fifteen types of help-seeking problematic situations that lead novice users to seek help. They also identified possible causes behind those situations, such as users' knowledge structure related to the domain of the search tasks, system knowledge, and information-retrieval knowledge. They showed that help-seeking situations are heavily influenced by users' personal and cognitive characteristics as well as the nature of tasks [305].

2.4 Information Searching Process

As Figure 2.2 shows, from the IIR perspective, a searcher develops an information need when faced with a real-world task. This need arises due to a gap in their knowledge state [85], an unresolved inconsistency in their mind (problematic state), or an Anomalous State of Knowledge (ASK) [27]. To resolve the problematic situation, the searcher initiates the information searching process to fulfill their information need to accomplish a specific task. The searcher begins interacting with an IR system and starts a search session by formulating their information need as a query. Once the query is submitted, the retrieval system retrieves results and presents them to the searcher in the form of a Search Engine Results Page (SERP). The searcher scrutinizes the content on the SERP and examines the title, link, and summary snippets for each document listed on the SERP for potential relevance to their information need. While examining these results, the searcher may find a particular summary interesting enough to warrant further investigation, prompting them to click on the provided link. This takes the searcher to the corresponding document, allowing for a more detailed examination to determine its relevance. If unsatisfied, the searcher may return to the SERP and continue exploring additional results. Throughout this process, the searcher continually learns about the task at hand, and the interaction cycle may prompt them to reformulate their query as their underlying cognitive focus and mental model of the task goal evolves. This, in turn, leads to a revised SERP that offers more promising results. This iterative process could go on until, eventually, the searcher decides to end their interactions, either within a particular SERP or the entire search session. The decision to stop may stem from fulfilling their information need, frustration with the retrieval system's inability to provide relevant results or various external factors such as time constraints.

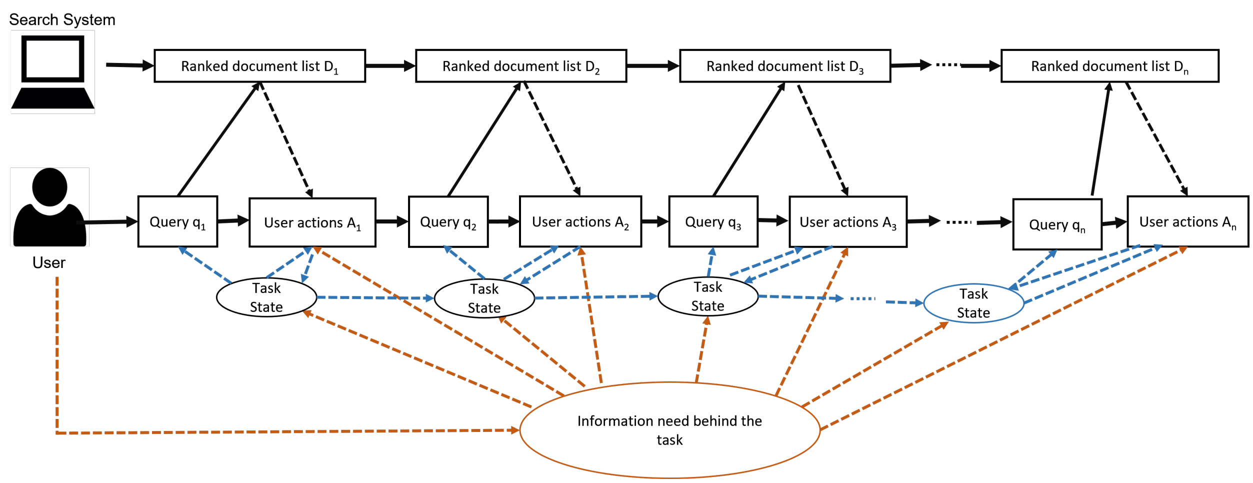

Figure 2.2: Information searching process in general from IIR perspective.

2.4.1 Search Session

The concept of a search session is fundamental in the search process, as it serves as a basis for developing metrics for application analytics and conducting behavioral analyses of user-centered systems [198]. In the context of IIR, a search session refers to a series of interactions between the searcher, the search system, and the retrieved document content within a specific timespan to address a single task or information need [118]. A search episode can contain one or more search sessions. A can contain one or more query episodes [3]. Sessions allow us to extend our understanding beyond individual query segments by preserving semantic connections between query trails and capturing the context of user activity. This enables a more comprehensive view of user interactions and helps to establish meaningful associations between related queries.

Previous IIR research has explored different strategies for identifying sessions from search logs [4]. For instance, content-based heuristics [264] utilize the lexical content of queries to determine shifts in the topics of query streams. Navigation-oriented heuristics [263] infer browsing patterns based on HTTP referrers and associated URLs. Time-oriented heuristics [76] involves setting a threshold of inactivity between logged activities to indicate the end of a session and the start of a new one. The underlying assumption is that if there is a sufficiently long break between a user's actions, it is likely that the user is no longer active, and the session is assumed to have ended. Catledge and Pitkow [61] found that the mean time between logged events was 9.3 minutes and added 1.5 standard deviations to establish a 25.5 minutes inactivity threshold. Over time, this threshold has been simplified to 30 minutes, which is the most commonly used approach for session identification [89, 263, 216] in IIR evaluation and experimental studies. Additionally, Radlinski and Joachims [222] employed a 30-minute timeout and query similarity measures to define sequences of similar queries, known as query chains.

2.4.2 User Search Behaviors, Search Strategies and Intents

In the past, research on IIR has concentrated on studying the actions of users during search sessions and query segments to obtain information relevant to their tasks. This information

[3] A query segment refers to a search session segment that starts from one issued query and ends at the next query [203].

[4] "Search log" is the record of events that happens when a user searches for something using a retrieval system. The system logs various actions during the search, like entering a query or clicking on a link.

has been used to improve the effectiveness of search systems. Research has shown that by analyzing explicit and implicit actions of users, such as query formulation, browsing patterns, and clicking-scrolling, it is possible to predict specific task details and even the tasks themselves.

Studies have looked into how people search for information based on the type of task and how they interact with search systems. During a search session, users perform various actions like creating and modifying queries, viewing result pages, clicking and scrolling through documents, reading individual documents, and going back to the search page to make changes. User behavior, such as how long they browse or how many times they click, can indicate their preferences. These actions can be categorized based on the query they are associated with [172]. While some studies examine behavioral signals computed over the entire search session (e.g.,[210]), others focus solely on signals available within a query segment (e.g.,[205]).

Marchionini categorized the activities of searchers during a search session into tactics, moves, and strategies [188]. The sequence of search tactics typically represents information search strategies employed within a search session. A search tactic, which is a segment of a search strategy, encompasses the actions performed within a query segment, including query formulation, search result browsing, and examination. Search actions, also known as moves, are fundamental components of search interactions and occur at various stages of the search process (e.g., formulating a query, examining the first search result, and continuously browsing the current search engine results page).

Expanding on this work, Xie proposed the concept of "interactive search intentions" and established a correlation between strategies and high-level intentions, such as locating a specific link and acquiring domain knowledge [304]. Broder [37] further classified web searches into three categories based on information-seeking and search intentions: navigational, transactional, and informational. Rha et al. [227, 226] demonstrated that differences in reformulation strategies could be associated with variations in the abovementioned intentions. They investigated how different types and states (satisfied or unsatisfied) of information seeking intentions lead to distinct query reformulation reasons and strategies. Building upon these findings, Mitsui et al. [210] showed that these intentions could be predicted at the query segment level using machine learning techniques with query browsing features as input.

2.4.3 Information Search Process Models

The search process is designed to capture the cognitive processes involved in a search session. In IIR research, various theoretical models have been developed to describe the general information seeking and search process. These abstracted theoretical search models have been widely utilized in IR research and implicitly integrated into different evaluation measures.

Prominent information search process models, such as those proposed by Ellis [90], Bates [17], Kuhlthau [147], Vakkari [276], and Järvelin [120], have primarily focused on the cognitive and affective aspects of the search process. These models emphasize the reciprocal influence between the searchers' seeking behavior, the search environment, the task context, and querying actions. Additionally, they highlight the socio-organizational context's impact on searchers, as influenced by their interpretation of the task and their understanding of the domain, ultimately affecting their behavior.

Vakkari [276] and Kuhlthau [147] have studied complex long-term tasks that consist of multiple smaller sub-tasks. While Kuhlthau [147] focused on affective and conceptual processes and developed the Information Search Process (ISP) model, which identifies six stages of information seeking (refer to Figure 2.3), Vakkari [276] translated this process into a series of stages from task assignment to task completion. Different stages of the search process correspond to users' varying cognitive focus, leading to different information sources and search actions. At the initial phase of the search process, searchers often experience uncertainty, with their cognitive focus shifting from frustration to clarity [147]. As they progress in

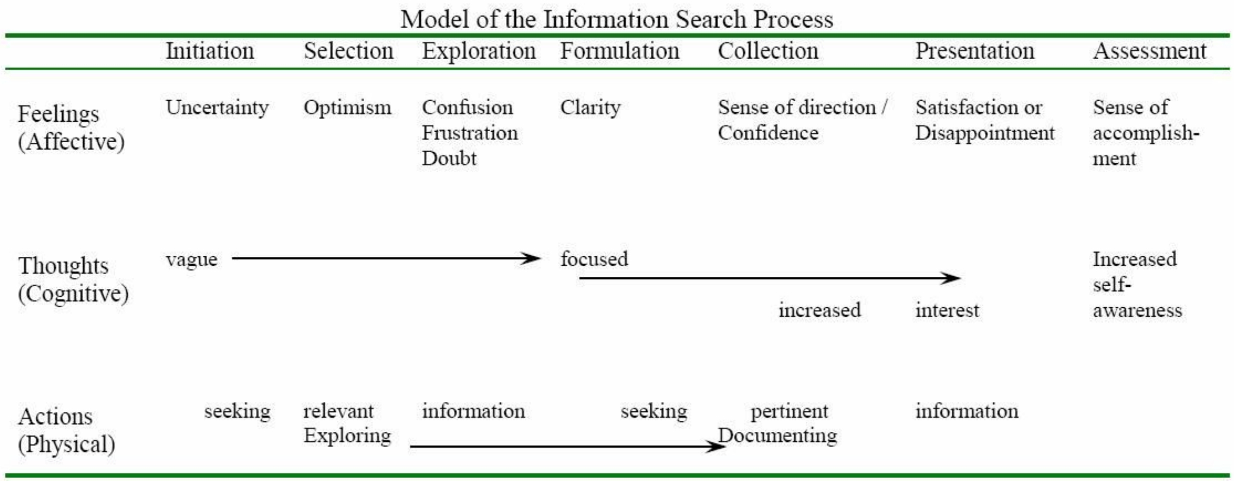

Figure 2.3: Information Search Process by Kuhlthau [147].

their search, their level of uncertainty diminishes [147]. The presence of uncertainty creates a chance for interactive systems to provide on-the-spot assistance and support for specific tasks. The ISP model's definition of uncertainty connects emotional and practical aspects of seeking information and is linked to various cognitive challenges in human-information interaction. This includes the anomalous state of knowledge during interactive information retrieval [20] and the cognitive gap in the Sense-Making process [85, 83].

Vakkari [276] expanded upon the previous model and introduced a task-based information retrieval process composed of three task stages: pre-focus, focus formulation, and post-focus. This extended model explores the connections between stages of the learning task process and information searching behaviors (refer to Figure 2.4). Vakkari's model studied further to demonstrate a strong correlation between the users' problem states during task execution and their information needs, the search strategies they employ, as well as their assessment of document relevance and usefulness.

Ellis [90] analyzed various ways people search for information, including browsing, filtering, and finding complementary information. This complements Kuhlthau's model by connecting to the different stages of the search process, including associated emotions (like

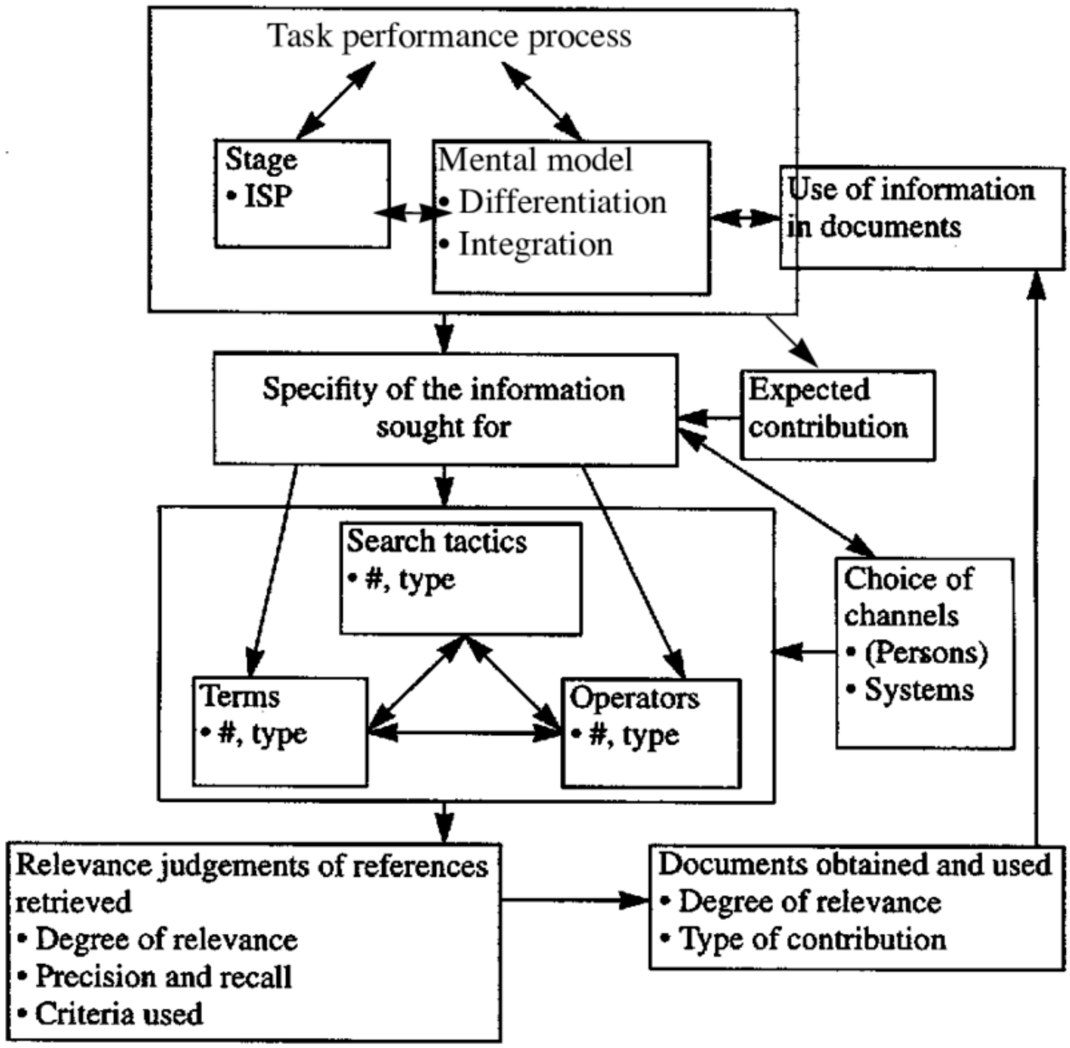

Figure 2.4: Task-based interactive information search process by Vakkari [276].

uncertainty or a sense of direction), cognitive states (like being vague or focused), and actions taken. [90].

Similarly, Bates [17] introduced the berry-picking model for searches, where the query evolves instead of having a single, static goal. This process is more like picking berries from bushes rather than following a linear sequence of steps to find the most useful documents.

Ingwersen [114] developed a framework aiming to emphasize the causal relationship between a searcher's information need, the problem state, the domain work task, and task interest (see Figure 2.5). Järvelin [120] later extended this model and added nine different task dimensions: the work task dimension, the search task dimension, the actor dimension, the perceived work task dimension, the perceived search task, the document dimension, the algorithmic search engine dimension, the algorithmic interface dimension, and the access and interaction dimension.

The models mentioned earlier are often used to describe the stages of task-based information seeking processes. However, they do not provide guidance on personalizing and optimizing search system support based on the searcher's evolving task stages and characteristics, such as their focus, intentions, search strategies, and problems encountered during the search. This means that most search processes and task-based models have limited practical implications for creating computational frameworks supporting complex tasks simultaneously. As searchers interact with new information, their knowledge state changes, leading to changes in their cognitive attention, information needs, and the usefulness of documents. Therefore, it is essential to consider the dynamic nature of task processes for search system results to benefit searchers in their current task state.

Researchers have recently created models that use advanced statistical and machine-learning techniques to represent task states or stages. These models are also computationally compatible with technical implementation requirements. For example, Rose and Levinstone [229] created a conceptual framework for user goals associated with queries in

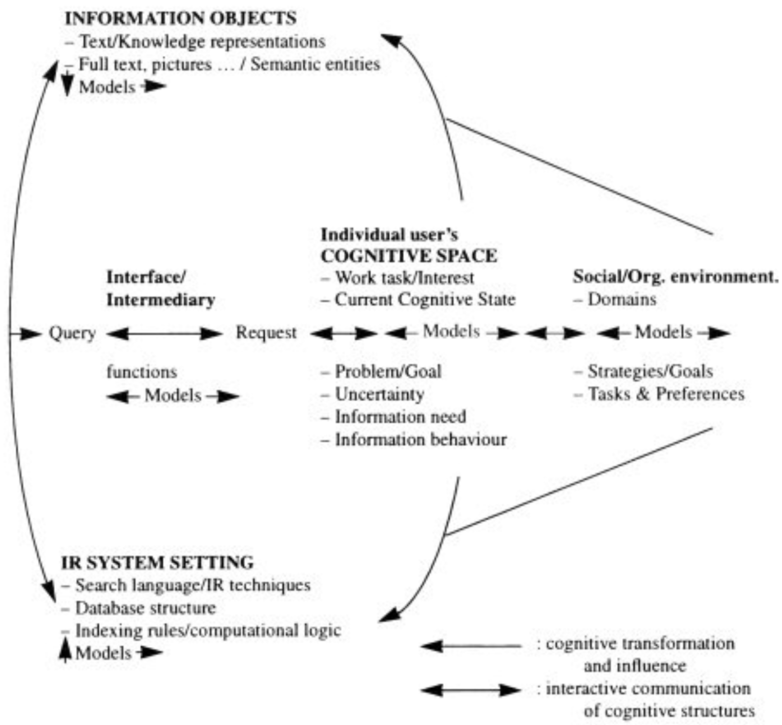

Figure 2.5: A task-based information search process defined by Ingwersen [114].

a way search systems can understand. Fuhr [97] and Dung and Fuhr created a framework using Hidden Markov Models to recognize and analyze search behaviors. Their approach includes both discrete and continuous signals and is a significant step towards building better interactive search systems. Cole [75] developed a framework that examined how users performed tasks of various kinds and discovered that the types of tasks and their difficulty levels could be identified by analyzing the user's activity and cognitive processes. This is done by examining the sequence and distribution of their activity and cognitive states. Luo et al. [181] applied a 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 to mathematically model dynamics in task-based session searches. They also simulated optimized recommendations based on the estimated states in the interaction process. Finally, Maxwell [192] represented a graphical and probabilistic iterative framework (see Figure 2.6) of the search process with attention to varying states at each step of the process, including the importance of background knowledge, intentions, and the searcher's subjective notions of relevance at each step of issuing queries and evaluating results.

However, further research is needed to connect the overall understanding of a task with its execution in search mechanisms.

2.5 Tasks in Information Seeking, Searching and the Process of Information Retrieval

Tasks are an integral part of people's daily lives, encompassing various activities, constraints, and durations for completion. The motivation behind information seeking and searching behaviors often arises from tasks that prompt individuals to transform their information needs into manageable goal-oriented subtasks and often engage in lengthy, iterative search processes characterized by distinct stages and evolving objectives. Consequently, understanding information seekers' goals, needs, and information-related issues involves considering the tasks

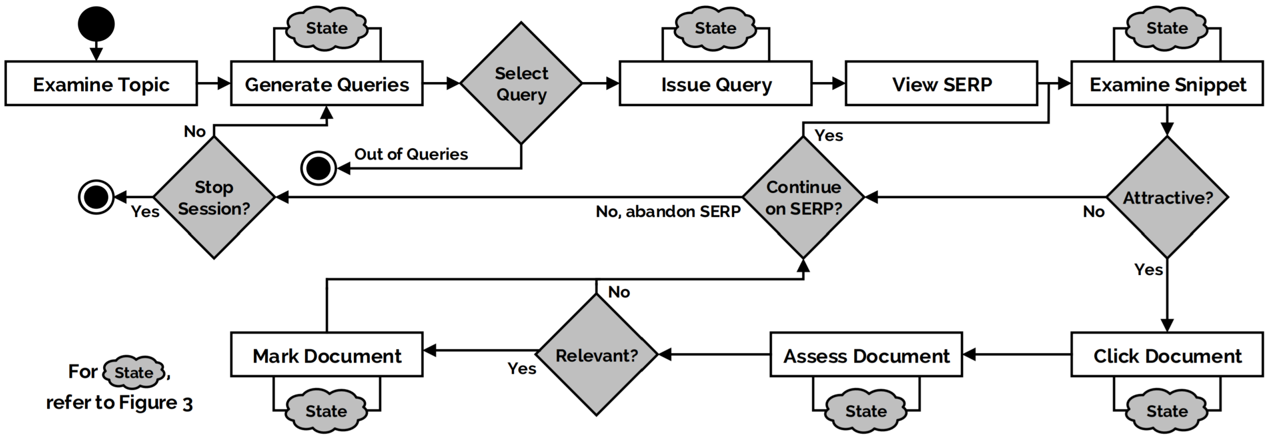

Figure 2: A model of the search process typically used in simulations. Here it has been augmented to show that various interactions invoke changes in the user's state - denoted by the *State* subprocess clouds.

Figure 2.6: Task-based interactive information search process by Maxwell et al. [192].

that drive their information searching behavior and the tasks they must accomplish during the search process [258]. Tasks significantly influence the challenges searchers face, their choice of information sources, their perception of success, and the difficulties encountered in finding useful and credible information [285, 242, 241].

In the field of IIR research, tasks serve two purposes. First, they are used to observe behavior and evaluate systems and tools [270]. Second, they can be the main focus of the research [270]. Research on tasks examines how their characteristics influence users' needs and behaviors. As an object of research, task characteristics can be experimentally compared as variables. For instance, tasks can be used to test whether a search system is effective for simple fact-finding tasks or problem-solving tasks. Tasks are just one aspect of research design and can be assigned to users in a user study to observe their behavior and/or system performance.

There are two main research methodologies used in task-based IIR: controlled user studies and large-scale search log analysis. Large-scale log analyses involve studying millions of queries and documents from real-world user interactions (e.g., [179]). This approach provides valuable insights into how people naturally search for information during real-life tasks. On the other hand, researchers focusing on task-based information seeking with real users often use user studies. These studies can occur in controlled laboratory settings where variables can be managed. Participants can work in their own space, and the task is an integral part of the research design. While modeling searchers' information needs in an experimental study can be challenging, task-based studies allow researchers to combine a specific task with a related topic to form an information need [137].

2.5.1 Defining Task

The task is a widely examined and essential aspect of individuals' information seeking behavior within the field of IIR [115]. According to cognitive and socio-cognitive perspectives, information seeking occurs within a task's context. The task context is primarily shaped by work or everyday life situations, serving as the central situational element. These situations give rise to problematic scenarios where individuals recognize their need for information to progress with their current task [115]. Within this definition, the concept of a task holds various interpretations. For instance, a task can be understood as an explicit expression of an individual's abstract and visceral information need [267] conveyed through spoken language. It also represents the overarching goal or purpose of the search process [137]. Such tasks include "gathering information to open a bank account" or "planning a trip." Therefore, a task fundamentally encompasses a series of interconnected physical, affective, or cognitive actions and activities that individuals undertake to achieve their goals, both in their professional and personal lives [277, 42]. From a more detailed perspective, a task can also be viewed as an atomic information need, leading to formulating one or more queries directed at IR systems [126].

2.5.2 Anatomy of Task

The classification and conceptual deconstruction of tasks in order to enhance search and IR systems have been extensively studied. These tasks possess inherent hierarchical and multidimensional characteristics and can be interpreted and modeled at various levels of granularity [45, 44]. Various disciplines have examined the influence of task characteristics on behavior and outcomes, providing insights into the complex nature of the “task” construct. For instance, early studies by Carter et al. [55] introduced an activity-based classification of tasks, encompassing intellectual construction tasks, mechanical assembly, and discussion tasks. In contrast, Hackman [101, 102] specifically focused on reasoning tasks, such as discussion tasks, production tasks, and problem-solving tasks. In HCI research, tasks are defined as goals to be achieved, with intentions serving as specific actions towards that goals (“To get something done, you have to start with some notion of what is wanted - the goal that is to be achieved” [213]). Norman [213] referred to these specific action statements as intentions, which transform goals into actionable steps. A *goal* is something to be achieved, often vaguely stated. An *intention* is a specific action taken to reach the goal. Patern‘o [217] also conceptualized tasks as multi-level entities.

In IIR research, Bystr”om [45] developed a nested task representation comprising three levels: work task, information seeking, and search or retrieval task. Each level encompasses its own set of goals, intentions, conditions, actions, and outcomes (refer to Figure 2.7). Furthermore, these task levels can be modeled as concrete sequences of states and actions [173]. Due to the influence of situational aspects during the search process, users’ task states often undergo changes and evolution throughout the search process [173].

At the outermost level, the concept of work task encompasses the overarching objective that initiates the search process [44, 191]. For instance, a task may involve the broad goal of “gaining knowledge about mesothelioma in order to comprehend available treatment options.” This overarching task consists of one or more information seeking tasks, such as “researching

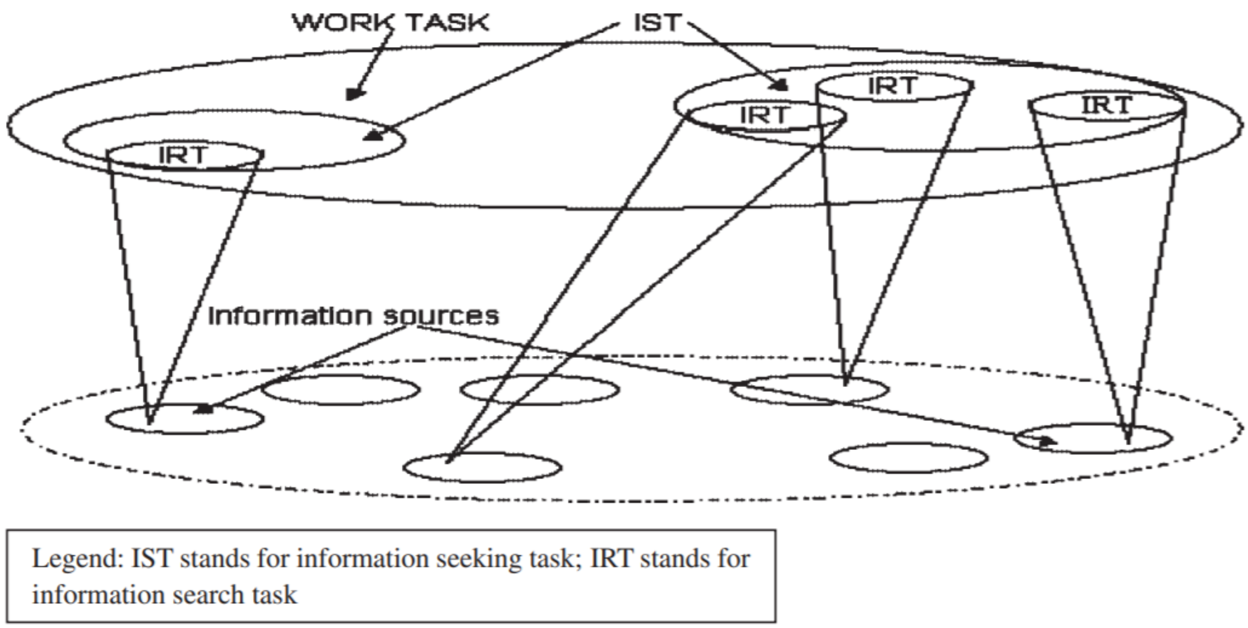

Figure 2.7: Deconstructing the task: A nested model with three task levels defined by Byström and Hansen [45].

the causes of mesothelioma" or "learning about different diagnostic tests." These tasks can be accomplished by consulting various information sources, including search systems and human experts. Additionally, each seeking task can be further broken down into micro-level information retrieval or search tasks, which can be fulfilled by consulting search systems in a single instance [44, 191]. Moreover, each information search or retrieval task typically represents a single or sequential atomic unit of information required, where a query is issued (e.g., "early symptoms of mesothelioma") and a series of associated actions, such as viewing and clicking on relevant results, are performed.

In a related study, Xie et al.[303] investigated the multi-level characteristics of searchers' objectives and proposed a hierarchical framework consisting of four levels of tasks. This framework, depicted in Figure2.8, incorporates different task attributes such as task goal, product, and process. The four levels encompass long-term goals, which represent searchers' personal interests; leading search goals or work tasks; current search goals, which encompass ongoing information seeking and search tasks; and interactive intentions, which encompass the specific objectives a searcher aims to achieve at each local stage of search tasks. Recently, Shah et al. [258] proposed another model of hierarchical task modeling that could be implemented in search and recommendation systems.

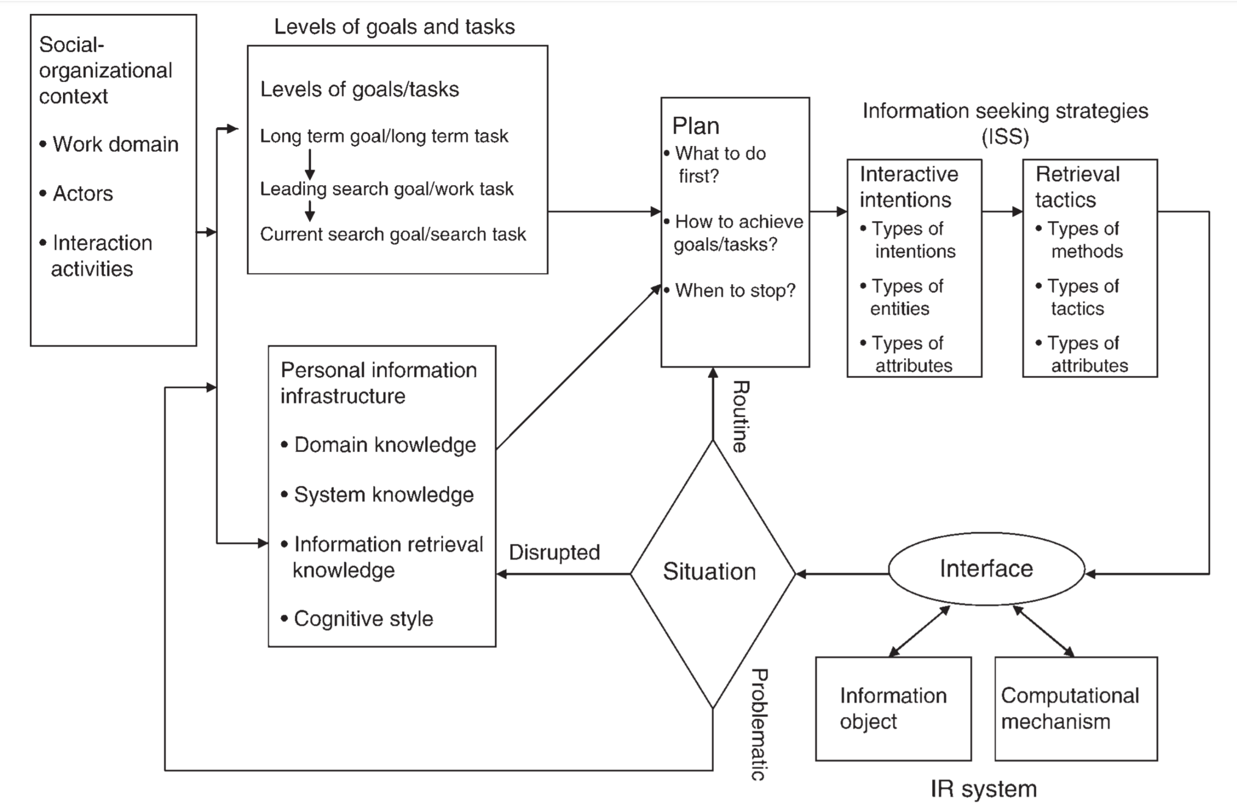

Figure 2.8: A hierarchical model of task goals defined by Xie [303].

In both classification schemes, each task level is characterized by a defined objective but an expected and potentially unknown outcome [270]. Moreover, each sub-task level relies recursively on the preceding one(s) [191]. Each task stage encompasses a series of physical, cognitive, and affective activities that represent the searcher's process of retrieving information from IR sources [276]. The specific type and quantity of activities a searcher performs in each task stage can vary based on the task's characteristics, such as its goal, product, difficulty, or complexity, as well as situational factors like time constraints [44]. For instance, in complex information domains like healthcare, searchers are required to engage in more cognitive activities such as reading, interpreting, and evaluating retrieved information, accessing multiple information types from various sources in an iterative process until a satisfactory solution is found or the task is abandoned [270].

The search process, which encompasses the journey from the primary search goal to the search outcome (i.e., the final task product), is composed of one or more sub-tasks that must be completed to achieve the ultimate task outcome [270]. Each task has a starting point that initiates the search process and a stopping point when the goal has been achieved or when the task is abandoned [270]. The triggering event may originate from a work context, a personal or situational context, or maybe prompted by another task [270]. Initially, the concept of a task was closely associated with work [276], where tasks were assumed to arise within a workplace context (e.g., "write a project report") [276]. However, Savolainen [245] argued that tasks could also emerge from everyday life interests, problems, or situations outside of a work context, leading to active information seeking and searching practices (e.g., "planning a vacation" or "finding a doctor"). Therefore, the term "work task" can be used to encompass both work-related and everyday life tasks [115]. It should be noted that specific components of large macro-level work tasks may require system and human support beyond the capabilities of search systems (e.g., "writing a dissertation proposal") [270, 258].

2.5.3 Task Categorizations

In IIR, tasks have been categorized in various ways to explore the relationship between task characteristics and behavior. In terms of task definition during the search process, previous research has developed and applied task models that consider task characteristics such as task complexity, task goal, and task outcome or product [187, 139, 141]. For instance, studies have combined task facets related to goals and products to define known-item tasks (with factual product/specific goal), known-subject tasks (with factual/amorphous goal), interpretive tasks (with intellectual product/specific goal), and exploratory tasks (with intellectual/amorphous goal) [156, 166, 74, 75, 63, 227, 206, 172]. Other commonly used task facets for modeling tasks include the level of document evaluation [75, 166] and the objective complexity of the task [75, 166]. Some of these task facets remain static throughout the task process. In contrast, others, particularly user-dependent attributes, may change during the search process, such as users' perceptions of the task, perceived task difficulty and complexity, knowledge about the task topic, and the intention behind each query [173].

When defining and representing tasks, multiple aspects are taken into account. These include the person carrying out the task, the task itself as a separate entity, and how the person performing the task views and approaches it. Existing research on task conceptualization and modeling can be broadly categorized into two groups: (1) *static representations of task*, which conceptualize a task as an abstract representation of predefined goals; and (2) *dynamic representations of task*, which model a task as a concrete sequence of states and actions [44]. Within the static perspective, tasks can be further classified into two categories: (i) *one-dimensional*, which defines a task based on a single attribute value such as task goal [138, 301] or task prior determinability [53, 51]; and (ii) *holistic/multi-dimensional*, which defines a task as a combination of multiple facet values [156, 166]. Thus, task properties can be divided into static task facets and dynamic task facets based on the task facets identified by Li et al. [156] and other relevant works [138]. Static factors include the task's type, goal, product, and objective complexity. Dynamic facets include things like how difficult a task seems, how much knowledge a searcher has about the topic, the searcher's intentions behind each query, their perceptions of the task, and their expectations for the search process and outcomes [173]. Figure 2.9 gives a holistic picture of all facets.

	Facets	Sub-facets	Values	Operational definitions
Generic facet of task	Source of task		Internal generated	A task motivated by a task doer. It is a self-motivated task
			Collaboration	A task motivated through discussion of a group of people
			External assigned	A task assigned by task setters based on their individual purpose
	Task doer		Individual	A task conducted by one task doer
			Individual in a group	A task assigned and completed by different group members separately, though they are in a group
			Group	A task conducted by a group of people (at least two people)
	Time	Frequency	Unique	A task conducted at the first time
			Intermittent	A task conducted more than one time but assessed by task doer as not frequently conducted
			Routine	A task assessed by task doer as frequently conducted
		Length	Short-term	A task which could be finished within a short time period (e.g. less than one month)
			Long-term	A task which has to be finished within a long time period (e.g. more than one month)
		Stage	Beginning	A task which just launched
			Middle	A task which has been running for a while and in the middle way
			Final	A task which is almost done or has been completed
	Product		Physical (for WT)	A task which produces a physical product
			Intellectual (for WT and ST)	A task which produces new ideas or findings
			Decision/Solu-tion (for WT)	A task which involves decision making or problem solving
			Factual information (for ST)	A task locating facts, data, or other similar information items in information systems
			Image (for ST)	A task locating images in information systems
			Mix product (for ST)	A task locating different types of information items in information systems
	Process		One-time task	A task accomplished through one process without repeated procedures
			Multi-time task	A task accomplished through repeatedly engaging in the same or similar process
	Goal	Quality	Specific goal	A task with explicit or concrete goals
			Amorphous goal	A task with abstract goals
			Mixed goal	A task with both concrete and abstract goals
		Quantity	Multi-goal	A task with two or more goals
			Single-goal	A task with only one goal
Common attributes of task	Task charac-teristics	Objective task complexity	High complexity	A task which involves significantly more paths during engaging in the task
			Moderate	A task which may involve a few paths but not significantly more during engaging in the task
			Low complexity	A task which involves a single path during engaging in the task
		Interdependence	High interdependence	A task conducted through collaboration of a group of people (at least two people)
			Moderate	A task conducted by one task doer with suggestions or help from other people or group members
			Low interdependence	A task conducted by one task doer without any help from other people
	User's perception of task	Salience of a task	High salience	A task assessed by the task doer as highly important
			Moderate	A task assessed by a task doer as moderate importance or the degree of salience of the task depends on specific situations
			Low salience	A task assessed by the task doer as unimportant
		Urgency	Immediate (urgent)	A task assessed by a task doer as highly urgent
			Moderate	A task assessed by the task doer as moderately urgent or the degree of urgency of the task depends on specific situations
			Delayed (not urgent)	A task assessed by the task doer as no urgency
		Difficulty	High difficulty	A task assessed by a task doer as high difficulty
			Moderate	A task assessed by a task doer as moderate difficulty or the degree of difficulty of the task depends on specific situations
			Low difficulty	A task assessed by a task doer as no difficulty or easy to complete
		Subjective task complexity	High complexity	A task assessed by a task doer as highly complex
			Moderate	A task assessed by a task doer as moderately complex or the degree of complexity of the task depends on specific situations
			Low complexity	A task assessed by a task doer as simple
		Knowledge of task topic	High knowledge	A task assessed by a task doer as highly knowledgeable on the task-related topic
			Moderate	A task assessed by a task doer as moderately knowledgeable on the task-related topic or the degree of knowledge on the task topic depends on specific situations
			Low knowledge	A task assessed by a task doer as unknowledgeable on the task-related topic
		Knowledge of task procedure	High knowledge	A task assessed by a task doer as highly knowledgeable on the method or procedures for completing the task
			Moderate	A task assessed by a task doer as moderately knowledgeable on the method or procedures to completing the task or the degree of knowledge on the method or procedures depends on specific situations
			Low knowledge	A task assessed by the task doer as not knowledgeable on the method or procedures for completing the task

Figure 2.9: A Comprehensive categorization of task facets by Li and Belkin [156].

Representing tasks based on various task facet values from static perspective

Theoretical support for a significant body of task-based IIR has been provided by both one-dimensional and multi-dimensional approaches in the task definition. The properties of the overarching task significantly impact multiple aspects of a sub-task [154], thus initiating the search process. While a multi-dimensional approach based on task goals and products was taken by Xie et al. [303], Bystr”om et al.[45, 44] adopted a one-dimensional approach to task classification, considering different levels of work tasks. Nevertheless, they defined hierarchical task classification schemes covering various levels of information-related tasks and sub-tasks. Apart from these studies, other researchers have defined the overarching task based on single or multiple combinations of task attributes. Campbell [49], based on subjective task complexity, created sixteen possible task types, which can be categorized into four groups: simple, decision, judgment, problem, and fuzzy tasks. Ingwersen [115], on the other hand, defined two types of tasks: work tasks and search tasks, based on the search process and goal.

Many studies have used comparable methods to model search and seeking tasks, resulting in significant redundancy. Marchionini [187] and Kelly [139], considering the task result or product, classified tasks into two types: close-ended and open-ended. Reid [224] categorized search tasks based on the source, distinguishing between internally and externally generated tasks. Broder [37] categorized Web search into three classes: navigational search, informational search, and transactional search. An informational task can be seen as a general topical search. In contrast, a navigational task corresponds to a known-item search. A transactional task is also a form of known-item task. Other works have examined search tasks such as fact-finding, information gathering, browsing, and transaction tasks [121, 250], or tasks that are either simple fact-finding or exploratory [6, 150]. Kim [141] identifies four task types that represent the diverse information needs of web searchers: factual, descriptive, instrumental, and exploratory. Exploratory tasks, among them, are often ill-defined, complex, and open-ended, requiring multiple rounds of searches. Various researchers have defined tasks as combinations of goals and products, resulting in multiple possible task types [181, 63, 228, 227, 166, 123, 3]. Figure 2.10 and Figure 2.11 depict different task classification schemes developed by these researchers.

Li and Belkin [156] have developed a comprehensive understanding of tasks by combining various works on task classification and their empirical studies on task-based information searching. Their faceted classification captures various generic task facets, including source, task doer, time, outcome, process, and goal, as well as common attributes such as interdependence, objective complexity, and searchers' perceptions of tasks such as salience, urgency, difficulty, subjective complexity, and knowledge. This scheme can be used to classify all levels of tasks and provides a more complete picture of the different dimensions of IR tasks than many previous studies, as shown in Figure 2.9.

In this dissertation, we utilize Li and Belkin [156]'s task typology to develop our user studies.

Modeling tasks based on various task facet values from dynamic perspective

Following Xie's concept of interactive intention in the hierarchical task model [303], the task state during a search task can be defined as the objectives or intentions that a searcher aims to accomplish within a query segment, which is a segment that starts from one issued query and ends at the subsequent query [211]. This encompasses a local goal and the associated activities that drive a query segment within the search task [172]. Searcher task states often undergo changes and evolve during search interactions, influenced by local situational factors. The dynamic task facets encompass attributes that combine the task and the searcher, which vary across different task states within the search process. These facets include subjective task complexity, difficulty, interactive intentions, strategies, task-related perceptions, and searchers' expectations.

Summary of work task types and dimensions

Author	Category/context	Task types	Task facets
Hackman (1968)	Intellective tasks	Production tasks Discussion tasks Problem-solving tasks	Content (what tasks are about); process (how tasks could be completed)
Whitley and Frost (1973)	A research laboratory (scientific tasks)	Responsibility tasks Extension tasks Development tasks Research tasks	Process
Tushman (1978)	R&D settings (the project's task)	Basic research Applied research Development Technical service	Task attribute (internal characteristics of tasks) (objective task complexity)
MacMullin and Taylor (1984)	Problems	Design/discovery Well structured/ill structured Complex/simple Specific/amorphous goals Initial state understood/ not understood Assumptions agreed upon/ not agreed upon Assumptions explicit/not agreed upon Familiarity/new pattern Magnitude of risk Susceptible/not susceptible to empirical analysis Internal/external imposition	Task attribute (structure; complexity; assumption of problems; familiarity; risk; analyzability); source of task (internal vs. external imposition)
McGrath (1984)	Group tasks	Generate (planning/ creativity) Choose (intellective/ decision-making) Negotiate (cognitive conflict/mixed-motive) Execute (contests / performances)	Process
Campbell (1988)	general tasks	Simple tasks Decision tasks Judgment tasks Problem tasks Fuzzy tasks	Task attribute (objective task complexity)
Byström and Järvelin (1995)	Municipal administrations	Automatic information processing task Normal information-processing tasks Normal decision tasks Known, genuine decision tasks Genuine decision tasks	Task attribute (perceived task complexity)
Algon (1997)	A drug development project team	Interaction with others Interaction with ideas or information Interaction with things	Process
Xie (1998)	Information intensive work tasks	Long-term goals Leading search goals Current search goals Interactive intentions	Goal (what tasks aim to), product (what tasks produce), and process
Byström and Hansen (2002)	Information intensive work tasks	Work tasks Information-seeking tasks Information retrieval tasks	Process, goal
Ingwersen and Järvelin (2005)	Work and daily-life tasks or interests	Work tasks Search tasks (retrieval tasks and seeking tasks)	Process, goal

Figure 2.10: A list of work task classification schemes summarized by Li and Belkin [156].

Summary of search task types and dimensions

Categories	Authors	Task types	Task facets
Search tasks	Marchionini (1989)	Closed tasks Open-ended tasks	Product
	Qiu (1993)	Specific tasks General tasks	Product
	Walker and Janes (1999)/Kim and Allen (2000)	Know-item search tasks Subject search tasks	Product
	Reid (2000)	Internally generated tasks Externally generated tasks	Source of task (from where tasks are created)
	Gross (1995, 2001)	Self-generated tasks Imposed search tasks	Source of task
	Kim (2006)	Factual tasks Descriptive tasks Instrumental tasks Exploratory tasks	Product
	Hert and Marchionini (1998)	Tasks based on pragmatic dimension Tasks based on semantic dimension Tasks based on syntactic dimension	Pragmatic, semantic, and syntactic
	Ingwersen and Järvelin (2005)	Routine search task Normal search task Genuine search task	Subjective task complexity

Figure 2.11: A list of search task classification schemes summarized by Li and Belkin [156].

The difficulty of a task is mainly determined by how the person doing it perceives the level of effort and difficulty involved. It is a subjective perspective [293]. Conversely, a task topic denotes the subject area that forms the focus of the task [137]. In contrast, domain knowledge or topic knowledge represents a person's general understanding of the task's subject area. The level of domain knowledge can influence a searcher's search strategies [204, 254].

Among the various attributes associated with tasks, task complexity has been acknowledged as a crucial factor affecting the information seeking strategies of searchers [41, 274]. Different studies have conceptualized task complexity from different perspectives [29, 240, 51]. The complexity of a task can either be objectively defined or perceived subjectively by those who are doing the task. Objective task complexity is not dependent on the person doing the task and can be determined in advance. Byström [47] viewed task complexity as prior knowledge about a task and expected outcomes. As task complexity increases, searchers require more sources of information, domain knowledge, and problem-solving information.

The difficulty level can depend on the task requirements and goals when searching for something. As the search progresses, the complexity can change based on how the searcher interacts with the search system. This cognitive effort and learning needed to complete the task is called subjective complexity. More complex tasks may require multiple searches, a higher cognitive load, and conscious decision-making. Additionally, in these tasks, the required information or expected results may not be known in advance [37].

2.5.4 Relationships Between Task Characteristics and User Behaviors in Search Sessions

Numerous studies have employed diverse contextual and behavioral data to examine the relationship between task characteristics and search behaviors. These studies have observed explicit and implicit search behaviors, such as dwell time on result pages, in order to infer searcher and task characteristics, including search satisfaction and task goals or products.

To explore the impact of tasks on search behaviors, researchers have intentionally manipulated task goals, creating tasks with specific, abstract, or mixed goals [63, 181, 228, 227]. They also have various task products involving tasks focused on locating facts, decision-making, or generating new insights from facts [142, 134, 187, 181, 63, 228, 227]. Task difficulty was found to affect search behavior differently depending on the task type (e.g., [167]). Additionally, the influence of task difficulty on searchers with varying expertise levels has been investigated, revealing differences in query term usage, query term richness, the proportion of terms derived from task descriptions, and search session (e.g., [163]). Task types have also been associated with variations in query length, session time, and URLs per query [166, 123]. Task complexity has shown similar relationships [138]. Tasks with a specific dimension resulted in a significantly more significant number of queries, more unique queries, longer query length, fewer clicks and bookmarks per query, and a smaller query log-likelihood [53].

Direct associations have been identified between the objective complexity of tasks and the time spent on tasks, time spent per selected document, and the total number of selected documents [155]. Liu et al. [165] examined task complexity and dwell time across task stages and discovered that document usefulness varied based on task complexity. By considering users ' expertise levels, Liu et al. [163] explored distinguishing task difficulty from behavior.

Similarly, Hienert et al. [110] found a significant relationship between searchers' actions and their familiarity with the topic, perceived difficulty, success, and time pressure. Differentiating between topic knowledge and domain knowledge has been accomplished through various behaviors, including content page and query browsing statistics [168], query reformulation patterns [112], the number of relevant information found [311], search efficiency [112], domain-specific word usage in queries [289], and differences in the types of web domains visited by experts compared to non-experts [289]. Task types have been distinguished based on various behavioral signals, including the number of content pages and queries, time spent on pages and queries, click and scroll depth, query reformulation behaviors, and task completion time [123, 166, 206, 139, 75]. Brennan [36] revealed significant direct relationships between task types (e.g., remember, analyze, create) and session length, the total number of queries, query length, and the total number of SERP clicks. Fox et al. [95] found found that click-through rates, time spent on search result pages, and how searchers exit result pages or end search sessions can predict user satisfaction. Furthermore, previous research has demonstrated that the usefulness of web pages can be implicitly measured based on users' dwell time on a page [290, 95] with a minimum dwell time of 30 seconds considered indicative of usefulness [254]. Web page relevance was also assessed using participants' bookmarks or snippets [254].

In the whole search session analysis, several significant search behavioral signals have been identified, including mouse activity and content viewing activity [6], time spent on content and queries [10, 100, 160, 161], total task duration [6, 169], bookmark statistics [6, 169], content page and query count statistics [6, 10, 100, 161, 163, 169], query reformulation behaviors [227]. Significant implicit and explicit behavioral signals in query segments include time spent on content pages and query pages [6, 160], mouse activities [6], viewing rank [6], and bookmark activities [6]. Important per-query segment and mid-session features include content page and query count statistics [169], page and query dwell time statistics [169], and query length statistics [10], and bookmark statistics [161].

2.5.5 Task Inference and Prediction from User Behavioral Signals

Many empirical studies have utilized a subset of the task taxonomy proposed by Li et al.[156] to infer task types based on search behaviors. These studies have employed task facets such as goals and products to define various task types. For instance, known-item search (factual product/specific goal), known-subject search (factual/amorphous goal), interpretive search (intellectual product/specific goal), and exploratory search (intellectual/amorphous goal) tasks have been formulated by combining task facets [74, 75, 206, 166, 123]. Task inference from behaviors has involved determining task goals and products. Several studies have attempted to predict tasks by utilizing browsing behaviors from entire sessions as features in a single predictive model [206, 123, 166]. In contrast, others have focused on specific query segments [8] for task inference [156].

In the domain of task inference, various statistical and machine-learning techniques have been employed. These methods, including independent t-test, the Mann-Whitney U test, ANOVA, linear modeling, the Kruskal-Wallis H test, classification, clustering, and regression, have primarily been used to detect significant behavioral differences among multiple task types [74, 135, 157, 166], between two task facets (e.g., factual/intellectual product) [166], among different levels of complexity [271, 36], topic familiarity [157], task difficulty [10, 74, 160, 163, 160], and between two levels of time pressure [162]. SVM and multilayer perceptron have also been utilized to predict task type, goal, product, and query segment intentions [204, 206, 116].

Several recent studies have utilized sequential features from search sessions to deduce tasks. For example, Cole et al. [75] created a graphical model of how the searchers' thinking changed throughout their search and determined each task based on the patterns in their cognitive state using a Markov model. Similarly, Kotzyba et al. [144] tried to infer tasks using task states as different page types to model sequential behavior in a session using the Markov and Hidden Markov models. Furthermore, Mitsui [208] created a graphical model for task characteristics, user characteristics, and behaviors to empirically studied them as a network of dependencies and predicted tasks.

2.5.6 Task Extraction from Large-scale Search Logs

There is a substantial body of research focused on segmenting and organizing large-scale search logs to establish coherent structures and identify users' tasks [143, 200, 104, 124]. These studies aim to predict session tasks based on users' query logs. However, extracting and predicting task features from large search logs pose challenges due to the messy nature of real-world event logs. Consequently, research on search log data has explored various methods and types of search behavioral signals to define and group tasks. For instance, researchers have utilized click-entropy, navigational clicks, total and unique URL domain clicks, query length, and post-click query behaviors.

Most of the studies on task extraction have focused on analyzing textual query contents and other query-related features to model tasks from search logs. Jones et al.[125] attempted to extract tasks and sub-tasks based on temporal differences, edit distances between queries, query co-occurrence features, and similarity between queries' search results. Radilinski et al. [223] extracted tasks by analyzing query reformulations and clicks using a random walk on the bipartite query-document click graph. Lucchese [178, 179, 180] leveraged session-based queries and proposed various clustering algorithms, along with an efficient heuristic algorithm, to extract tasks from a given query collection. Lucchese [179] created a heterogeneous graph that links similar topics of searches with appropriate web pages, using the user's clicks as a basis. Wang et al. [283] employed a structured learning approach to partition tasks based on latent query-based, URL-based, and session-based features. Blei et al. [31]

adopted a topical similarity approach to cluster queries by assessing similarities between query topics, thereby identifying tasks. It is assumed that queries submitted within a short time are related to the same task, and users with similar information needs tend to submit topically coherent queries.

In contrast, Mehrotra et al. [199, 200] suggested two Bayesian non-parametric methods that utilize a hierarchical structures to extract subtasks from complex tasks. These methods recursively derive hierarchies of tasks and subtasks in the search process. On the other hand, Li et al.[152] employed topic modeling using Latent Dirichlet Allocation (LDA) combined with temporal grouping from Hawkes processes to group temporally close queries on the same topic into tasks. Additionally, several studies have explored word embeddings to determine tasks and subtasks [199, 200].

2.6 Various Approaches to Evaluate Task-based Search Process

With many task inference, extraction, and prediction models available that use different combinations of task and searcher characteristics, behavioral signals, and methods, it can be challenging to evaluate them. The main questions are how to assess the models and which metrics to use. Evaluation is the *systematic determination of merit and significance of something using criteria against a set of standards* [127]. Evaluating the effectiveness of new information retrieval systems and their components before implementation is a critical aspect of IIR research. The evaluation aims to assess the quality of the obtained outcomes, and the specific study designs and effectiveness measures vary depending on the research domain and problems. Different IIR evaluation techniques are used based on the nature of the retrieved results and quality measurement. When evaluating task-based IIR models, the ultimate goal is to determine the subjective or objective benefits for task performance [25]. Ingwersen [115] defined IR evaluation as nested frameworks of IR context, seeking context, work-task context, and socio-organizational and cultural contexts (Figure 2.12).

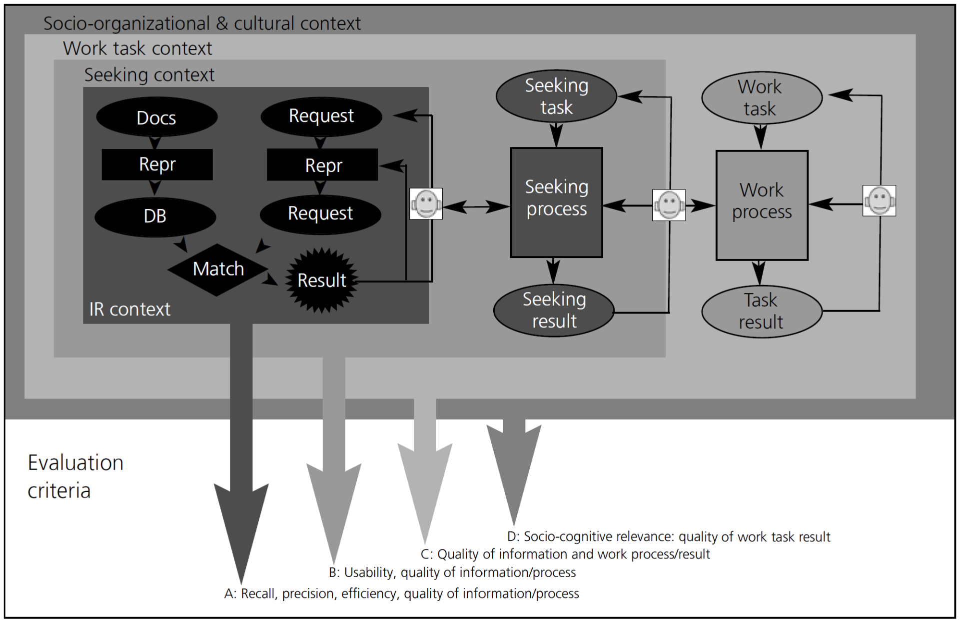

Figure 2.12: A nested evaluation framework by Ingwersen and Järvelin [113].

Evaluation criteria vary across different contexts. Various mechanisms for evaluation have been mentioned in the existing literature, including assessing model performance with gold-standard datasets or ground truth labeled data, conducting user studies, employing alternative evaluation techniques, and utilizing TREC Tracks [5] that are relevant to the problem. The test-collection approach is commonly used to gauge the quality and efficiency of retrieved results, employing metrics such as Recall, Precision, Mean Average Precision (MAP) [77], Discounted Cumulative Gain (DCG), and Normalized DCG [221]. A test collection consists of a database of documents that can be retrieved, a group of test questions or topics that represent information needs, and a recall base that identifies relevant documents within the database. Test-collection-based experiments do not involve direct participation from searchers. On the other hand, user-centered evaluations involve task-based studies where human searchers interact with a given IR system within a controlled task context. The

[5]https://trec.nist.gov/tracks.html

IIR communities also use TREC Tracks [6] for evaluation purposes, including the Session Track and Task Track. These tracks contain curated search log data from various topics and subject domains that have been manually annotated with relevance scores by expert assessors. Researchers can use these datasets to validate their predictive models and compare their performance with similar models. In a laboratory setting, such datasets can be used to evaluate the usability and users' satisfaction under controlled conditions in the seeking context.

To evaluate the performance of a model using a gold-standard dataset, finding or developing ground truth data is crucial. For instance, Lucchese [179] created a ground truth dataset by partitioning long-term sampled session data from actual query logs using a predefined time threshold and identifying distinct time-gap sessions. Within each time-gap session, human annotators manually grouped queries related to tasks. This dataset and evaluation metrics, such as F-measure, Pairwise Precision, or Recall, provide options for evaluating a task prediction model. Analyses of logs, longitudinal task-based evaluations [277], and evaluations based on interaction simulation [13] are becoming more prevalent today. Furthermore, a few studies have utilized experimental design to investigate the relationship between task performance and information searching (e.g., [298]).

A valuable way to evaluate is through user studies. Researchers can obtain feedback from people through platforms like Amazon Mechanical Turk to confirm the accuracy, usefulness, and relevance of predicted tasks. They can also gather sub-task labels for a selection of randomly assigned queries. As noted by [201], there are some alternate evaluation techniques, such as analyzing predicted task categories using qualitative methods. Kelly [137] categorized the IIR evaluation methods (Figure 2.13) into eight types, which can be grouped into three broad categories: an evaluation based on test-collections with no human interaction, user-centered laboratory-based evaluations, and user-centered operational evaluations in natural

[6] https://trec.nist.gov/data.html

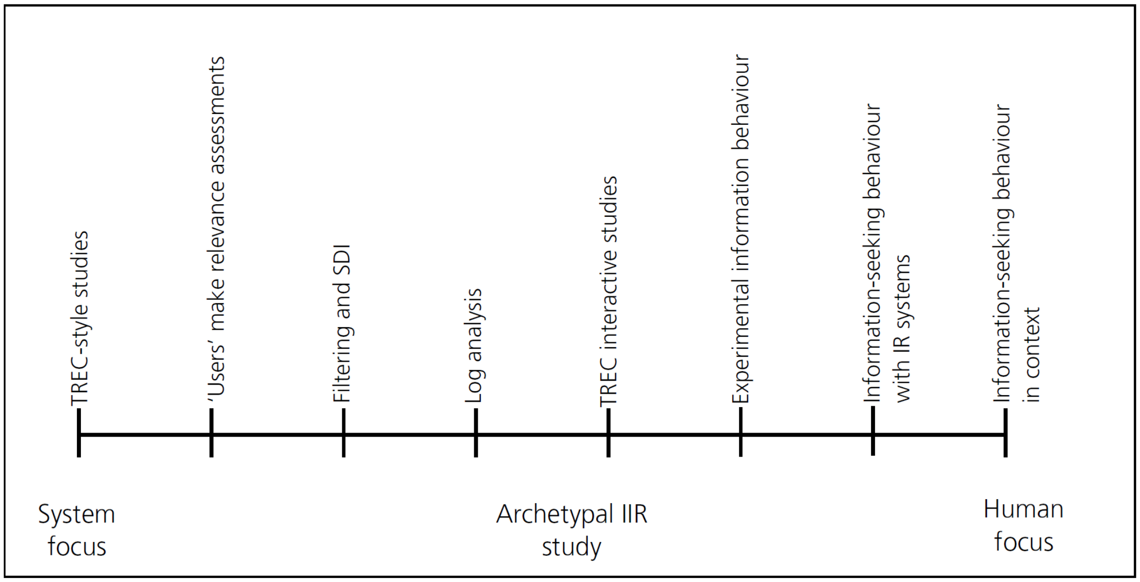

Figure 2.13: IR evaluation methods by Kelly [137]

settings.

Chapter Summary

- In this dissertation, the literature review provides context for tasks related to seeking, searching, and analyzing information at different levels and in various aspects.
- The review covers both static and dynamic properties of tasks, including search process models.
- The review summarizes recent research on complex search tasks and discusses the analysis of search behavior, including commonly used constructs such as tactics, strategies, and paths and the associated measures.

Part II

METHODOLOGY AND MODELING

In the second part of the thesis, we introduce the task representation model that provides a more realistic, conceptual model of the task-based information searching process. We also detail a series of user studies and experiments we conduct as part of our general methodology.

Chapter 3

METHODOLOGY OVERVIEW

This dissertation uses a mixed-method design [79] to take advantage of both quantitative and qualitative research approaches and develop a comprehensive study design. In a mixed-methods design, multiple research approaches are employed to collect and analyze data, bringing complementary perspectives to the research design and data collection and analysis. After considering the need for qualitative and quantitative research designs, data collection, and analysis procedures (sequential, concurrent, or transformative) for the dissertation, we have decided to use the explanatory sequential mixed method design in multiple research phases [79].

3.0.1 Research and Overview of the Procedure

Figure 4.1 presents a general overview of the research procedure.

3.1 The Role of the Researcher

Our roles and involvement as researchers are different at each for each study. First, we are "outside observers attempting to discover a law of relationships among observable features" [262] instead of directly involved in the research procedure. Thus, we primarily focus on administering a survey and a controlled user study, collecting data, and performing data analyses to investigate the characteristics of variables and the relationships between these variables. Later, we actively analyze large search log datasets to extract and model tasks. We are mindful of the biases in those datasets and algorithmic approaches and take necessary caution while making assumptions. We are involved in "a sustained and intensive

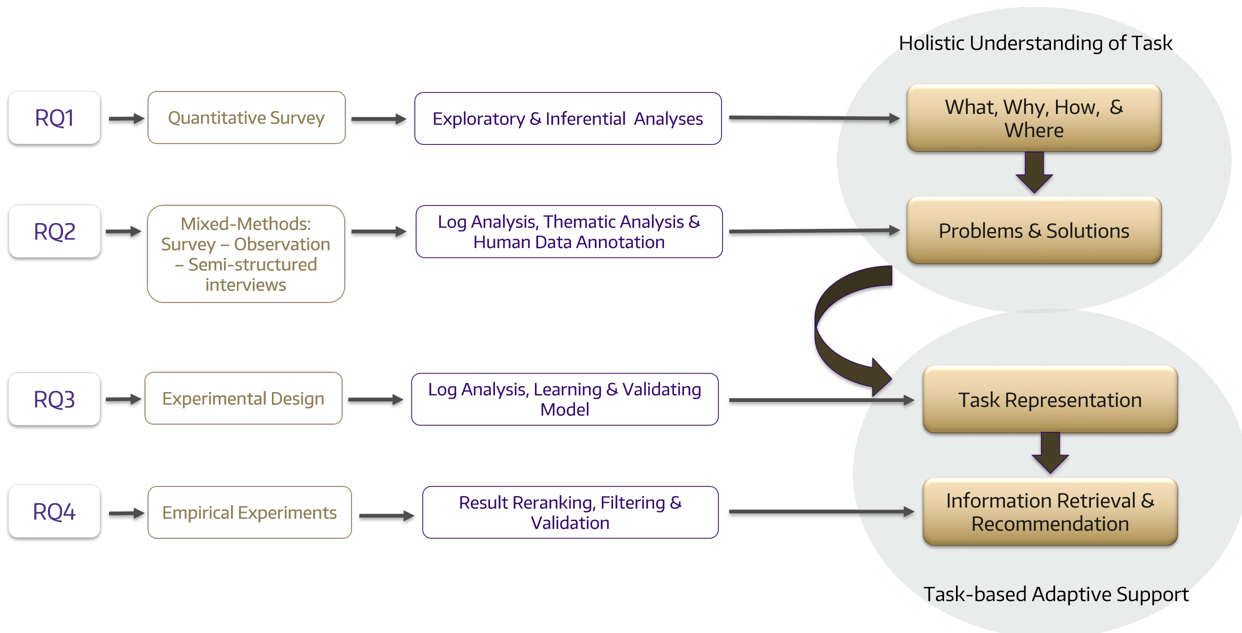

Figure 3.1: The outline of the research procedure.

experience with participants" [79], and also focus on "discovering the meanings constructed by the participants as they attempt to make sense of the circumstances they both encounter and create" [262].

3.2 Research Permission and Ethical Issues

Ethical issues have been considered at each step of this dissertation. In order to work with human participants, the research plans were reviewed by the Institutional Review Board (IRB) for approval. To protect participants' rights during data collection, an informed consent form was developed and presented to them before engaging in the study [79]. The information collected for this dissertation includes personal or identifiable data such as the participants' names, email addresses (for communication purposes), and audio transcripts of interviews. However, no questions or data analysis solicited personal or potentially harmful information. This personally identifiable information was kept confidential. Moreover, we removed identifiable information before performing any data analysis step. Each participant has a random identifier assigned to them that has been used to analyze and report the data associated with them.

Chapter Summary

- The dissertation uses a mixed-methodology based approach to address four research questions.
- Ethical issues have been considered at each phase of this dissertation.

Chapter 4

RESEARCH QUESTION 1: STUDY 1: A HOLISTIC NOTION OF TASK

The task is a crucial aspect of information seeking and searching processes, traditionally considered the initial phase of an individual's behaviors when seeking/searching for information. However, extracting task information from users can be challenging due to its elusive and often hidden nature. One approach to gain insights into how individuals perceive and express tasks is to observe their seeking and searching behaviors, focusing on their situated performative actions.

4.1 Introduction

Various behaviors and concepts are involved in information seeking tasks, including the recognition of an information need, the selection of information type, and the identification of information sources. In supporting users in complex search tasks [87], four essential elements for designing task-based intelligent systems can be identified. These elements include (1) representations of users' tasks, which refer to how users express their tasks and how systems comprehend and interpret them; (2) information, which pertains to the units of information retrieved in response to users' information needs; (3) representation or form of information, such as documents or web pages; and (4) interaction, which encompasses the structuring and presentation of information to users, as well as how users interact with the system and the retrieved information. These elements collectively aim to facilitate the exploration, sense-making, and acquisition of knowledge.

Only a few studies have taken a holistic approach to the seeking process, focusing primarily on specific elements (e.g., [47, 114, 275, 278]). Historically, most studies have either concentrated on one component or explored only a few aspects of a single element, neglecting the utilization of multiple sources and channels within a single information seeking process [147]. However, individuals tend to select their information strategy and sources based on the information needs associated with a particular task [241]. Therefore, an information seeking task process involves interconnected constructs, encompassing why people seek information, what information they seek, where and how they seek it, and which sources and channels they utilize. Therefore, it is a holistic process [256]. However, information seekers do not make decisions about why, how, what, and where to seek information by considering these constructs in isolation. Instead, information seeking is a holistic process where seekers make decisions influenced by all possible constructs and various dimensions of each construct [257]. Individuals develop and adapt their information-seeking strategies based on continually shifting values across several dimensions. These dimensions collectively form the users' cognitive focus, particularly in Chapter 6.

As an initial step towards task-based intelligent search assistance and to gain further insight into the role played by the constructs mentioned above in people's general information seeking strategies, the first study of this dissertation research involves conducting an exploratory study that examines the interactions among various aspects of the information seeking process.

In the first user study (Study 1) of this dissertation, we adopt a novel methodological perspective rooted in previous task-based studies on information interaction and Dervin's Sense-Making theory [84, 85, 83]. This study decomposes the information seeking task process into components such as information needs, sources, methods of accessing sources, representation of information (e.g., knowledge), and the interpretation or form of those representations (e.g., fact, advice, opinion). Our main objective is to explore how different tasks elicit specific information needs, which in turn influence the methods, strategies, and sources individuals employ when seeking different types of information.

Therefore, to address the high-level question posed in Chapter **??**, the following local research question guided our study design, data collection, and analysis techniques.

> **RQ1.** How different types of tasks trigger specific information needs that can lead to different methods and strategies for seeking different forms of information and information sources?
>
> **RQ1.1**: What is the relationship between information forms, information needs, information sources, and how are the sources accessed for tasks?

4.2 Study 1: Method

In order to comprehend the correlations between diverse information needs, forms of information, and information sources, an exploratory survey is carried out to investigate individuals' information seeking behavior in their daily lives on Amazon Mechanical Turk [1]. The survey is designed to explore how different types of tasks can elicit various information needs, which are triggered by fundamental human needs such as social, emotional, and cognitive aspects. These information needs, in turn, may result in different approaches to seeking information from a variety of sources. Thus, an exploratory user study is conducted to delve into this matter. In essence, by obtaining answers to the "what" question (regarding the nature and form of information) and the "why" question (regarding the information need), this study aims to predict the response to the "where" question (concerning the information source) within the information seeking process.

[1] https://www.mturk.com/

4.2.1 Survey Design

The survey participants were asked to imagine a situation where their decision-making or actions were significantly influenced by their process of searching for information [132, 66, 313, 215]. The survey comprised 15 distinct information scenarios, each representing a specific information requirement and the corresponding type of information needed to fulfill that requirement. Drawing on the categorizations presented in prior works such as [132, 66, 313, 215], the information needs were classified into four categories: cognitive, personal, social, and affective. Cognitive needs involve finding relevant information within one's immediate surroundings, society, and the world; seeking advice or opinions to aid decision-making; acquiring self-education knowledge; and gaining a sense of security through knowledge. Personal integrative needs entail finding support for one's own values, gaining insights into one's own life, and empathizing with others' problems. Social integrative needs include feeling a sense of belonging, establishing a basis for conversation and social interaction, and fostering a connection with other people. Affective needs encompass seeking social and emotional support for personal issues, seeking such support for others (e.g., family and friends), and searching for validation of personal thoughts or ideas. In accordance with theories proposed by Marchionini [190], Buckland [39], and Rulke [232], information representation was categorized into three types: object, knowledge, and social or sensory stimuli contextual. Additionally, the study drew on various forms of information from Shah et al. [255], which were consolidated into four broad categories: advice, opinion, fact, and social or emotional support. By combining these different aspects, a total of 15 unique combinations were obtained, resulting in the creation of 15 distinct scenarios. Figure 4.1 presents and outline of the survey scenario development process.

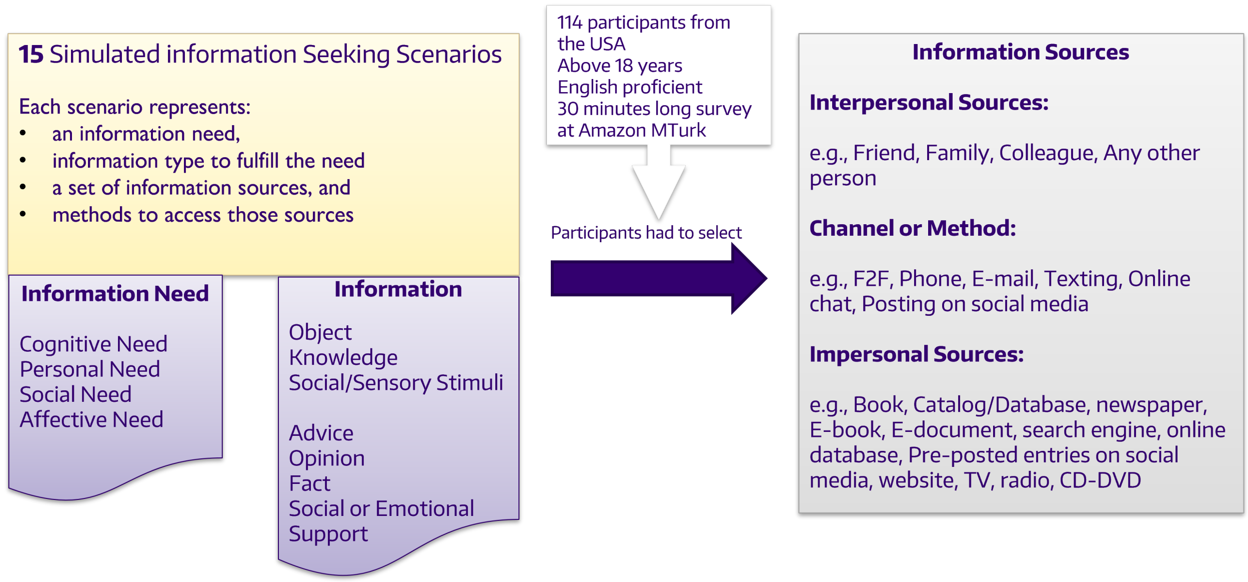

Figure 4.1: The outline of the survey scenario development procedure.

Developing Scenario

The 15 task scenarios encompassed various combinations of three dimensions, representing different aspects of information seeking: information needs and information itself. These scenarios were constructed based on Borlund's concept of simulated task scenarios, as mentioned in [32]. Borlund emphasized the importance of realism in these scenarios, ensuring they reflected participants' real-life situations. To achieve this, we derived the scenarios from ordinary everyday life situations.

Each scenario presented two sets of information sources. The first set consisted of human or interpersonal sources and the methods to access them. The second set included all impersonal sources, encompassing online and offline information retrieval systems. The methods to access these sources were implied in their names, hence not explicitly mentioned. The assignment of scenarios to participants did not follow any specific order. A detailed description of each of the 15 scenarios is provided in the following section.

To ensure the study's validity and account for the participants' background influence on their responses, we deliberately decided not to collect demographic or personal data from the participants. The aim was to understand source preferences in a general context, independent of specific contextual factors. However, the lack of demographic information and the artificiality of the task scenarios, coupled with the unknown demographics and actual information needs of the MTurk workers who participated, might raise questions about the survey responses' contextual relevance. To address these concerns, we sought the expertise of two researchers specializing in interactive information seeking and retrieval. These experts were consulted during the scenario creation process to ensure that each scenario accurately represented cognitive needs, personal needs, social, affect-driven needs, and various forms of information. Additionally, we conducted rigorous pilot surveys and evaluations involving five participants from diverse backgrounds. This iterative process allowed us to gather feedback on scenario descriptions, language, and the participants' ability to relate to the task described in each scenario.

Scenarios for the Survey

Scenario 1: Cognitive need, Informative object, Fact Suppose you and your friend went to the grocery store together to buy produce. You are discussing the attributes of different vegetables and fruits; taste, nutrition, etc. Your friend is explaining that the skin of "Red Delicious" apples is thinner than the skin of "Granny Smith" apples. You are not fully convinced of this and you want to find out whether it is true. Where and how do you look for information?

Scenario 2: Cognitive need, Knowledge, Fat Suppose you are just about to start your first semester. You have decided to buy a new laptop instead of taking your 5-year-old computer to the college. There are different types of laptops with various specifications and you cannot decide which laptop you would buy a Mac or Windows? Would you prefer a 4 GB Ram or an 8 GB Ram? You have decided to look for some facts on how to compare different laptops. Where and how do you look for information?

Scenario 3: Cognitive need, Knowledge, Opinion As a student, suppose you have to create a poster presentation for your social science class. You have done presentations with PowerPoint but have never presented with a poster. Since this is your first poster, you are most concerned about how such a document should look. You want to find some opinions concerning the best ways to organize a poster layout so that you are prepared to create your project. Where and how do you look for information?

Scenario 4: Cognitive need, Knowledge, Advice This summer, you are traveling to Australia. You have never been there before and there are many tourist attractions you would like to visit. Therefore, you want information to help you plan your journey. You have set aside one month for the trip and hope to see as much of the country as you can. As you are unfamiliar with the territory, you would like to know about the places before going there. Where and how do you look for information?

Scenario 5: Personal need, Social or Sensory stimuli, Social or emotional support Suppose you have a -3.0 power in both eyes and you wear eyeglasses regularly. In your recent visit to your ophthalmologist, your doctor prescribed contact lenses for clearer vision. You have never used contact lenses before, so you would like to know about overall experience of wearing them - are they easy to wear or painful? Based on your findings you would like to decide whether to wear the lenses. Where and how do you look for information?

Scenario 6: Personal need, Social or Sensory stimuli, Fact Suppose you recently read a book on health, genetic disorders, and diet, and it has influenced the way you think about your health and lifestyle. You have learned that there are some diseases like cancer or diabetes can be caused by the combination of mutations of inherited genes, lifestyle choices, and your environment. Many rare diseases and conditions usually develop when an individual is born with a mutated gene. After reading the book, you want to know if a rare disease or common health conditions run in your family so that you can make precautionary changes in your existing diet or lifestyle. Where and how do you look for information?

Scenario 7: Personal need, Knowledge, Opinion You have followed the news and debate about legalizing marijuana in certain US states. In your personal opinion, alcohol and cigarette smoking are much more dangerous than marijuana.Now you are curious to know how others feel about this issue. Where and how do you look for information?

Scenario 8: Personal need, Knowledge, Fact Suppose your father is returning home tomorrow from the hospital after a prolonged illness. Tomorrow also happens to be his 60th birthday. On this special occasion, you want to make your grandmother's old chicken broth recipe to surprise your father.However, you do not know the recipe and you need to find it out. Where and how do you look for information?

Scenario 9: Social need, Knowledge, Fact You will be attending asocial gathering this evening. It is a birthday party for a friend being held at a local restaurant. You do not know many of the guests in attendance. You thought you could facilitate conversations with new people if you were up-to-date on some recent topics of interest. You have decided to look into a wide expanse of events since you do not know the other guests' interests and backgrounds. Where and how do you look for information?

Scenario 10: Social need, Social or Sensory stimuli, Social or emotional support Your cousin, a senior in college, said that one of her friends started to smoke. You fear your cousin might begin smoking in the near future and decide to educate her about the risks, so you have to find some information on what could happen if she starts smoking. Where and how do you look for information?

Scenario 11: Affective need, Informative object, Social or emotional support Imagine your family members are out of town for a family wedding in California. You were unable to go because of a work commitment. It is now a few days since they have gone and you are missing them very much, especially when you are at home all alone. You are feeling very sad and in order to feel better you would like to find something to cheer you up. Where and how do you look for it?

Scenario 12: Social need, Social or sensory stimuli, Advice A friend of yours has an appointment to get a flu shot tomorrow, but she has a cold. She is debating whether to keep her appointment. She asked for your advice. You would like to help her by finding out what is generally recommended for people in her situation. Where and how do you look for information?

Scenario 13: Affective need, Knowledge, opinion Suppose you have inherited a large sum of money left by your recently deceased uncle. You are not sure about what to do with this money. You are considering investing it in the stock market as a bond or corporate stocks. However, you are unaware of stock market trends and lack the knowledge required to make a sound judgment on what to do with your inheritance. You would like to find information to help you pick the best type of financial instrument for the investment. Where and how do you look for information?

Scenario 14: Affective need, Knowledge, Fact After graduating as a veterinary doctor, you are just about to start a career in the veterinary industry. You are concerned about how to work towards your retirement.Therefore, you want to know what you should expect from your industry. Where and how do you look for information?

Scenario 15: Affective need, Knowledge, Social or sensory stimuli Your best friend is getting married and asked you to be in their wedding party. You have been friends since the 5th grade. You feel touched, excited, and also a little stressed. You have to plan the shower and bachelor(ette) and you want them to both be great and memorable for your friend. You have no prior experience and no idea where to begin. Therefore, you need information on how to become a perfect head of the wedding party. Where and how do you look for information?

4.2.2 Data Collection

As mentioned above, we conducted the survey was conducted on Amazon's Mechanical Turk (MTurk), which allows individuals or businesses to post intelligent human tasks for users from around the world. The demographically diverse participants on MTurk have been found to provide reliable and high-quality data, similar to other conventional survey collection methods such as standard Internet samples and typical American college samples [40, 177].

Since this study was exploratory in nature, we decided that using the participant pool on MTurk was an excellent choice due to its accessibility and size. To ensure the data collected was reliable and high-quality, we took several measures, such as recruiting expert participants, offering better compensation rates, and keeping the tasks manageable. Moreover, we also included attention-check questions throughout the survey to assess the participants' attentiveness and the randomness of their responses.

We collected responses from 114 MTurk workers who were at least 18 years old and from the USA. The workers were presented with a 30-minute questionnaire with 15 scenarios where they had to choose the information sources and search methods from an exhaustive list. Participants could select any source from either impersonal or interpersonal categories for each scenario. Every participant completed all 15 scenarios. Typically, most respondents took 15 to 30 minutes to finish the survey. We downloaded and quantitatively analyzed the recorded responses, resulting in a final dataset comprising participants' source selections in 1,710 scenarios.

4.3 Data Analysis

The research employs an exploratory approach to examine the survey responses (refer to Tables 4.3 and 4.1). The descriptive statistics indicate that most respondents chose to utilize a search engine for scenarios 1, 2, 3, and 4. The second-largest group of participants consulted with individuals, primarily friends and professionals. Most respondents preferred interpersonal sources over impersonal ones regarding scenarios centered around personal needs, factual information, opinions, and social support (scenarios 5, 6, 7, and 8). In scenario 5, friends and search engines were equally popular choices (80 and 83 times, respectively). At the same time, family members were preferred in scenarios 6 and 8, and friends in scenario 7. Specifically, participants selected family members 87 times, friends 76 times, and family members again 111 times for scenarios 6, 7, and 8, respectively. The second most popular choice for these three scenarios was the search engine, selected 59, 69, and 43 times, respectively. Interestingly, there was a significant gap between these scenarios' most popular and second-most popular source choices. Most respondents consulted impersonal sources in socially-driven scenarios (9, 10, and 12). The popular choices for these scenarios included consulting friends (scenario 9), searching directly on professional websites, reading newspapers and news-media sites, and pre-posted entries on social media. For scenarios based on affective needs (11, 13, 14, and 15), respondents predominantly favored impersonal sources, particularly web searches, except for scenario 11, where interpersonal sources (friends) were preferred over other options. It is worth noting that a combination of impersonal and interpersonal sources was also observed, such as pre-posted entries on social media and colleagues, among others. Analyzing people's preferences for accessing interpersonal sources revealed an apparent inclination towards direct face-to-face conversations with familiar sources like friends and family members. Conversely, when seeking assistance from strangers (e.g., customer services, dealers), individuals often preferred indirect communication methods such as writing messages or posting online (see Table 4.2).

4.4 Implications and Discussion

Through the data analysis, specific patterns in information needs, sources, and methods have been identified. Some of these patterns are novel, while others support existing research. First, The increasing popularity of online sources, particularly pre-posted entries on social

Table 4.1: Participant's Use of Impersonal Sources

Scenario	Book	Offline Catalog	News-paper	Any Object	E-copies	Online Data-base	Google Scholar	Online Catalog	Social media Pre-posted Forum Entries	Websites	News media sites	Search Engine	TV channels	Radio
1	10	3	2	0	12	16	11	4	15	5	2	92	0	0
2	0	6	7	0	7	15	4	1	32	35	1	92	0	0
3	13	2	1	6	11	8	5	3	24	14	1	86	0	0
4	18	4	8	8	3	9	0	2	47	59	7	97	0	1
5	3	1	3	2	4	2	1	0	51	17	4	83	0	0
6	5	2	4	5	16	7	6	4	15	20	2	59	0	0
7	3	1	11	1	10	1	4	1	67	8	26	69	3	2
8	13	2	1	7	3	2	0	2	10	4	0	43	0	0
9	1	0	31	3	4	2	0	1	38	3	67	59	25	6
10	10	0	11	5	36	5	10	0	25	29	13	84	1	0
11	6	0	6	1	1	1	1	0	26	5	5	39	31	4
12	1	2	3	4	16	3	5	1	20	34	7	98	0	0
13	10	2	9	2	11	2	2	0	29	56	13	77	2	0
14	4	1	7	1	11	9	1	0	37	44	3	90	0	0
15	11	2	14	1	1	2	0	0	68	16	3	100	2	0
Total	108 (6.32%)	28 (1.64%)	118 (6.78%)	46 (2.69%)	146 (8.54%)	84 (4.91%)	50 (2.92%)	19 (1.11%)	504 (29.47%)	349 (20.42%)	154 (9.01%)	1168 (68.30%)	64 (3.74%)	13 (0.76%)

Table 4.2: Participant's Use of Channels to access Interpersonal Sources

Methods	Friend	Family	Colleague	Stranger
Face to Face	282	253	144	98
Phone	108	184	15	22
Texting	132	97	22	2
Online Chatting	59	18	4	25
Email	15	16	18	10
Posting on online forums	9	4	5	144
Mail/Letter	0	0	0	0

Table 4.3: Participant's Use of Interpersonal Sources

Scenario	Friend	Family	Colleague	Stranger
1	23	20	2	26
2	60	33	21	31
3	50	18	48	11
4	28	24	5	33
5	80	54	21	20
6	9	87	2	13
7	76	42	21	18
8	3	111	0	2
9	49	17	10	10
10	17	19	2	18
11	84	37	2	2
12	19	16	5	27
13	23	32	15	47
14	12	8	48	29
15	72	54	6	14
Total	605 (35.38%)	572 (33.45%)	208 (12.16%)	301 (17.60%)

media and online forums for engaging with strangers, highlights the evolving nature of human information-seeking behavior. A growing preference for online sources indicates a rise in collective and collaborative information seeking. Second, participants commonly relied on both friends and family as their primary sources of information. However, they also used all four types of interpersonal sources in every scenario, except in scenario 8, where "family" was the only choice. However, their utilization of impersonal sources was limited to a few types, such as web search engines, pre-posted entries on social media, professional websites, online news media sites, electronic copies of journals and magazines, as well as newspapers and books/manuals. Third, while participants were required to choose at least one source, the majority preferred to use multiple impersonal and interpersonal sources across all scenarios. This suggests that individuals generally prefer not to rely on a single method or source when seeking information. When searching for information, people usually use different sources in a circular process rather than a linear one. However, when it comes to finding a family recipe, people usually rely on personal connections to obtain the information they need. This was the case for the task of finding a grandmother's chicken broth recipe, where interpersonal sources were exclusively used. Fourth, although web search engines are favored due to their accessibility and ability to provide a vast amount of information quickly, the study reveals that people also rely on other sources depending on their specific information needs. The choice of using a book, newspaper, or other physical sources is determined by the nature of the required information. Thus, the motivations and expectations of users significantly influence their selection of information sources.

This study has found that people choose their information sources and strategies based on their specific needs for a particular task. This aligns with previous studies (e.g., [241, 286, 242]) that show individuals rely on others for critical activities and seek assistance or opinions. The study also highlights that people prefer multiple sources, both impersonal and interpersonal. The findings emphasize the importance of learning in the information-seeking process and offer practical design considerations for facilitating users' interaction with information.

Despite increasing interest in this task-based IIR, there is still a lack of task-based search and intelligent assistance. This study aims to bridge this gap by adopting a holistic perspective on the information seeking process, connecting individualized and social search processes with the searcher's context and task.

Chapter Summary

- The study's overall objective is exploratory and to observe information seeking based on the previously identified four aspects of information seeking behavior and various dimensions of them.
- The study explores how different types of information and information needs may lead to different sources and channels of seeking information.
- Popularity of online sources indicates a shift towards collective and collaborative information seeking behavior.
- Individuals prefer to use multiple sources, both impersonal and interpersonal, indicating a non-linear and repetitive information seeking process.
- Individuals often select sources based on the task characteristics and their specific needs (interpersonal for personal/intimate scenarios, impersonal for cognitive needs).
- Individual motivations and expectations behind tasks significantly influence the choice of information sources.

- When people decide to use a particular source, their decisions are heavily influenced by the type of information they are looking for and the kind of information needs they have.

Chapter 5

RESEARCH QUESTION 2: STUDY 2: CHARACTERIZE TASK, POTENTIAL SEARCH PROBLEMS AND SOLUTIONS

In the previous chapter 4, we have examined components of information seeking and searching processes that define users' seeking/searching tasks. However, past empirical works show that users often fail to achieve their intended goal at the end of their search processes [285]. Moreover, existing IIR studies have shown that individuals frequently face difficulties seeking information (e.g., [246]). These include internal barriers (e.g., lack of knowledge, unable to articulate and express the need) [28], external barriers (e.g., time constraint, institutional restrictions) [246], and interpersonal barriers (e.g., lack of help from other people) [266]. Therefore, encountering and solving problems are also parts of users' journey toward task completion.

5.1 Introduction

In this second study, we explore users' perceived obstacles and desired assistance at different stages of information searching. We analyze a range of implicit and explicit user search behaviors to gain a deeper understanding of the task as a comprehensive entity. Consequently, we can concentrate on representing a user's undisclosed task objectives in relation to those explicit behavioral cues. Our approach involves soliciting explicit feedback from information seekers regarding the barriers or difficulties they encounter and the assistance they seek at that particular moment. We assume a strong correlation between their task state and these two aspects of information seeking [82, 85, 17, 128, 72]. With this in mind, we investigate the following research questions (RQs) in connection to the overarching research question

RQ2 presented in Chapter 1, which establishes a link between the study of behavioral signals and problems and assistance.

RQ2 What problem(s) do searchers encounter while performing a specific task, and the type(s) of help do they prefer to get from the system?

RQ2.1: To what extent can the problems or barriers perceived by an individual in different stages of information search be predicted from different search behaviors?

RQ2.2: To what extent can the best help for an individual be predicted from the search behaviors and the perceived barriers/problems at a given moment in the information search process?

RQ.2.3 What problem(s) users encounter while performing a specific task type?

RQ.2.4 What type(s) of help they prefer to get from the system to resolve those problems?

RQ.2.5 What is the distribution of problems and preferred help across tasks of different types?

To address these research questions, we explore various characteristics of tasks, information seekers' implicit behaviors, and explicit feedback in the course of online information search episodes, including the problem(s) they face and suggestions for help(s) to address solutions to the potential difficulties in finding information.

5.2 Study 2: Methodology

We devised a controlled laboratory-based user study within a research university setting, employing a mixed-method approach. The study involved participants who were instructed to undertake three search tasks. In addition to completing the tasks, participants were prompted to document any challenges they encountered at different stages of the search process and express their requirements for assistance at those particular moments. The subsequent subsections outline the vital elements of the study procedure.

5.2.1 Participants

We recruited 37 participants from a United States research university's undergraduate student population by advertising on email lists, Facebook groups, and around campus. Participants registered online and were compensated with $20 in cash once they completed the study. The participants had diverse educational backgrounds, ranging from History and Arts to Computer Science, Psychology, Public Health, and more, and their ages ranged from 18 to 22 years, with an average age of 20.2. 70% of the participants were female, and all were native English speakers. Additionally, most participants reported over ten years of web search experience.

5.2.2 Search Task Scenarios

During their sessions in the laboratory, participants engaged in a sequence of three tasks. The sequence began with a warm-up task, followed by task 1 and task 2 in that specific order. Detailed descriptions of each task are provided below. Before undertaking each task, participants were required to complete a pre-task questionnaire regarding their familiarity with the upcoming task. Each task had a designated time limit of 5 minutes, 20 minutes, and 20 minutes respectively. However, participants were permitted to finish earlier if they desired. The tasks involved bookmarking web pages and generating reports based on their findings. The warm-up task served the additional purpose of acquainting participants with the system and the laboratory environment.

Our task design incorporated the manipulation of complexity level, task goal, and task product, drawing inspiration from the task classification scheme proposed by Li and Belkin [156]. We recognized that several characteristics identified by Li and Belkin [156] were implied in our task design. Given the nature of search scenarios and the study's design, it was crucial for the tasks to be performed individually ("Task doer" facet), possess uniqueness ("Time-Frequency" facet), be externally assigned ("Source of task" facet), and be completed within a short duration ("Time-Length" facet). Furthermore, we carefully selected task product, goal, and objective task complexity elements. As discussed in the previous section, we also included questions regarding participants' perception of task difficulty [29, 156, 301]. Additionally, the tasks encompassed different information needs, and these binary labels were utilized for subsequent task-based analysis. We tailored the tasks to ensure that the task situations were relevant and realistic for the study participants, particularly for the specific group involved (i.e., undergraduate students), based on the simulated work tasks of [33]. Here are the task descriptions:

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

Task 1 [Complexity = High; Need = Cognitive; Product = Intellectual; Goal = Amorphous] (20 min): *Lara Dutta of India was crowned Miss Universe in 2000, and between 1994 and 2000 women from India won two Miss Universe competitions, four Miss World competitions, and many less well-known competitions. These facts inspired you to explore the relationship between these wins and the Indian government's decisions and policies in your final paper for Indian Society class. To what extent can decisions and policies of the*

Indian government be credited with these wins? As a part of your final paper, please offer your brief answer to this question and identify the useful pages (from the bookmarked pages) which were actually used for constructing your answer.

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

5.2.3 Study Design and Data Collection

During the study, we collected the following information from the participants.

User Characteristics and States of Knowledge

Participants had to complete three search tasks. Before and after each task, every participant was requested to complete a questionnaire to gather their knowledge and skills about the task, including task dimensions and requirements like time and task type. Before searching, we asked participants about their knowledge of the task features in a pre-task questionnaire. After completing the task, we asked for feedback on their overall experience and their level of understanding of the task in a post-task questionnaire. Another important aspect is users' perceived difficulty of the search task. Understanding a task can affect users' search process and interactions with the system [29]. We also asked about their perception of each task's complexity. Therefore, the questions concerning task difficulty were adapted from [301].

Information Seeking Barriers and Help

At the beginning of each query segment, participants were given a questionnaire and instructed to open a new tab whenever they wanted to formulate a new query. Opening a new tab (Google search home page) triggered a pop-up window with a problem-help questionnaire, automatically presented by the browser. In this questionnaire, participants were asked to report the problems they had just encountered and specify the types of help they would like to have in a real-time search scenario. This design allowed us to gather information about real-time search problems and preferred forms of assistance before users constructed specific queries. To ensure valuable and accurate responses regarding problems and helps, we adapted relevant questions and concepts from previous research [287, 248, 67, 305]. The pop-up window included a list of problem and help questions, which can be found in Tables 5.1-5.2. By combining and consolidating various problems and helps identified in the studies, as mentioned earlier, we defined eight problem types and four help types that shed light on the problems encountered, strategies used, and patterns of interactions employed to overcome these difficulties. It is important to note that the problems and preferred helps defined in the typologies are not mutually exclusive, allowing participants to select multiple options if applicable. However, if they indicated ”no problem encountered” or ”no help needed,” they could not select any other option simultaneously, as those options were not applicable in such situations. The order of options for all questions presented in the pop-up windows was randomized to mitigate potential ordering effects.

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.

Table 5.1: The potential problems, and the possible responses. Acronyms in parentheses are used in data analysis section.

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

Table 5.2: The potential helps, and the possible responses.

Search Session Logs: Implicit and Explicit Behavioral Signals

We collected search logs for each task with several features regarding the tasks, users, and behaviors described in Tables 5.4-5.5. The behaviors that participants exhibit right before reporting a problem or seeking help can serve as indicators of their difficulty in finding information at that time. For example, suppose a participant did not use bookmarks. In that case, it may suggest that they encountered useless search results and would prefer web pages with more helpful information. This segment of behavior is directly related to the current search query and occurs during a specific time and space. Session-based features, on the other hand, represent longer-term behavior patterns. Current behaviors that occur after reporting a problem or seeking help may be a result of the issues encountered during the search process.

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

Table 5.3: Behavioral features for the current query segment.

5.2.4 Study Procedure

Participants were invited to the lab, where they conducted web searches using Google Chrome on a desktop computer that had the necessary applications installed. They were first pro-

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

Table 5.4: Behavioral features collected for the past query segment.

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

Table 5.5: Behavioral features accumulated over the session.

vided with a description of the study procedure, browser interface features, and a customized Chrome extension. Subsequently, participants performed three search tasks, which involved finding web information based on simulated task scenarios. Whenever they opened a new tab to formulate a query, the Chrome extension provided the above problem-help questionnaire. Participants were asked to complete short questionnaires before and after each search task, assessing their familiarity with the task, experience, perceived difficulty, and familiarity with the topic. During the study sessions, various aspects of participants' interactions, such as web page visits, bookmarks, queries, and problem-help answers, were captured using a custom Chrome extension and Morae [1]. Once all the tasks were completed, participants' sessions concluded with a brief semi-structured exit interview. The interview covered their satisfaction with the search sessions, their evaluation of the search process, problems they encountered, desired help, perceptions of their everyday search experiences, and their expectations of an ideal information retrieval system's performance. Each participant's laboratory session ranged from 50 to 80 minutes in total.

5.3 Data Analysis and Results

The data obtained from the user study underwent analysis utilizing descriptive and inferential methods. We focused on the problem and help situations occurring during specific task states, mainly when participants were about to create a query. We measured users' cognitive focus or task state during individual query segments to facilitate implementation. The problem-help situation was operationally defined as the occurrence when a user engaged in an information searching task encountered an obstacle that hindered further progress. Participants marked the identified problem(s) and desired help(s) in the questionnaire, formulated a query, and the situation concluded when they proceeded to plan or reformulate another query, moving forward in the search task. While the query segments were logged in

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

the system file, we manually ensured accurate identification of the boundaries for each unit of analysis by carefully reviewing and coding the dataset. Following this cleaning process, we conducted various analyses based on the problem/help annotations gathered from 273 query segments.

5.3.1 Descriptive Analysis

To examine the frequency of various problem-help scenarios, descriptive analyses were carried out. The participants selected a total of 378 problems, with the most common issue being the uncertainty about accessing comprehensive and valuable sources of information, which occurred 82 times. When encountering problems during the search process, the participants expressed a need for assistance, with a cumulative total of 253 instances. The preferred form of assistance, indicated by 86 participants, was receiving recommendations from the system regarding potentially helpful web pages.

The query reformulation strategies of participants and their associations with identified problems and helps during a specific query segment were examined by annotating each query according to the categories proposed by Rha et al. [226]: Generalization (1), Specialization (2), Word Substitution (3), Repeat (4), New (5), Spelling Correction (6), and Stem Identical (7). The most commonly employed reformulation type in our study was "New." Participants frequently created new queries using entirely new terms compared to their previous queries [226]. The second most frequently utilized reformulation type was "Word Substitution," wherein participants often replaced a portion of the query with new terms while retaining the length and some previous terms from the previous query. Notably, reformulation patterns varied between tasks. In both tasks 1 and 2, participants employed "New" query reformulations to initiate a new search. For task 1, they employed various types of formulation multiple times. The distribution of query reformulation types for the entire sample is depicted in Figures 5.1 and 5.2.

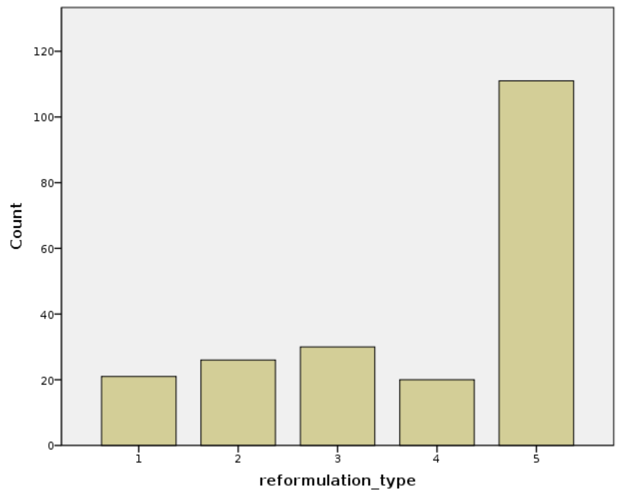

Figure 5.1: Counts of query reformulations - (1) Generalization, (2) Specialization, (3) Word Substitution, (4) Repeat, (5) New

Problem-Task Mapping

Figures 5.3, 5.4, 5.5 present the types of problems that participants selected for each task type. From the analysis, it is clear that task types have influenced the problems that participants experienced during the search process. From the figures, we can also see that task complexity level increased their problems. Most participants noted multiple problems at a given time, highlighting the necessity of understanding users' constraints from the systems' perspectives.

Help-Task Mapping

Similarly, Figures 5.6, 5.7, 5.8, which present the types of help that participants selected for each task type and problem categories, show that users need different types of support to address diverse problems that they encounter during search processes. Often they need to consult human sources to gain information on their tasks. Although these findings are not

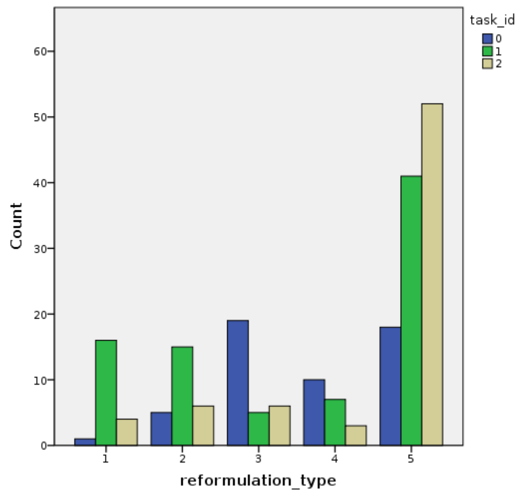

Figure 5.2: Counts of query reformulations throughout in each task session - (1) Generalization, (2) Specialization, (3) Word Substitution, (4) Repeat, (5) New for the warm-up (0), Task 1 (1), Task (2)

entirely new and have been observed in previous studies (e.g., [287]), they emphasize the limitations of query and document recommendations of current search and IR systems.

Task-Problem-Help Mapping

To examine the distribution of problems and preferred help across tasks of different types, we created an adjacency matrix for task type, problems, and preferred help. Figures 5.9, 5.10, 5.11 present the task:problem:help matrix for three task types that resulted from the preliminary examination of the data.

5.3.2 Inferential Analysis and Results

The next sub-sections provide the outcomes of the data analysis related to the research inquiries proposed.

task_id: 0

		difficult_articulate		irrelevant_results		topknowledge_lack		patient_lack	credibility_uncertain		sources_unaware		toomuch_information		source_unavailable	
		0	1	0	1	0	1	0	0	1	0	1	0	1	0	1
		Count	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count
Problem group	7	0	2	2	0	0	2	2	2	0	0	2	2	0	2	0
	8	0	1	1	0	0	1	1	1	0	1	0	0	1	1	0
	16	0	1	0	1	1	0	1	1	0	0	1	1	0	1	0
	20	0	1	0	1	1	0	1	1	0	1	0	1	0	0	1
	23	0	1	1	0	1	0	1	1	0	1	0	1	0	1	0
	24	1	0	1	0	0	1	1	0	1	0	1	0	1	1	0
	25	1	0	1	0	0	1	1	0	1	0	1	1	0	1	0
	28	2	0	2	0	0	2	2	2	0	0	2	0	2	2	0
	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	30	2	0	0	2	0	2	2	2	0	0	2	2	0	2	0
	31	4	0	4	0	0	4	4	4	0	0	4	4	0	4	0
	32	1	0	0	1	0	1	1	1	0	1	0	0	1	0	1
	35	7	0	7	0	0	7	7	7	0	7	0	7	0	7	0
	36	1	0	1	0	1	0	1	0	1	0	1	0	1	1	0
	43	1	0	1	0	1	0	1	1	0	0	1	0	1	1	0
	45	1	0	0	1	1	0	1	1	0	0	1	1	0	1	0
	46	7	0	7	0	7	0	7	7	0	0	7	7	0	7	0
	48	3	0	3	0	3	0	3	3	0	3	0	0	3	3	0
	50	1	0	1	0	1	0	1	1	0	1	0	1	0	0	1
	51	18	0	18	0	18	0	18	18	0	18	0	18	0	18	0
	52	4	0	0	4	4	0	4	4	0	4	0	4	0	4	0

Figure 5.3: Different types of problem combinations for task type 0 [Complexity = Moderate; Need = Cognitive; Product = Intellectual; Goal = Amorphous].

task_id: 1

		difficult_articulate		irrelevant_results		topknowledge_lack		patient_lack	credibility_uncertain		sources_unaware		toomuch_information		source_unavailable	
		0 Count	1 Count	0 Count	1 Count	0 Count	1 Count	0 Count	0 Count	1 Count	0 Count	1 Count	0 Count	1 Count	0 Count	1 Count
Problem_group	1	0	4	0	4	0	4	4	0	4	0	4	0	4	0	4
	2	0	2	2	0	0	2	2	0	2	0	2	0	2	2	0
	3	0	1	1	0	0	1	1	0	1	0	1	1	0	1	0
	4	0	4	0	4	0	4	4	4	0	0	4	4	0	0	4
	5	0	1	1	0	0	1	1	1	0	0	1	1	0	0	1
	6	0	4	0	4	0	4	4	4	0	0	4	4	0	4	0
	7	0	3	3	0	0	3	3	3	0	0	3	3	0	3	0
	9	0	2	0	2	0	2	2	2	0	2	0	2	0	0	2
	10	0	1	1	0	0	1	1	1	0	1	0	1	0	0	1
	11	0	1	1	0	0	1	1	1	0	1	0	1	0	1	0
	13	0	1	0	1	1	0	1	0	1	1	0	1	0	1	0
	14	0	1	0	1	1	0	1	1	0	0	1	1	0	0	1
	15	0	2	2	0	2	0	2	2	0	0	2	2	0	0	2
	16	0	2	0	2	2	0	2	2	0	0	2	2	0	2	0
	17	0	3	3	0	3	0	3	3	0	0	3	3	0	3	0
	18	0	1	1	0	1	0	1	1	0	1	0	0	1	0	1
	19	0	1	1	0	1	0	1	1	0	1	0	0	1	1	0
	20	0	3	0	3	3	0	3	3	0	3	0	3	0	0	3
	21	0	1	1	0	1	0	1	1	0	1	0	1	0	0	1
	22	0	1	0	1	1	0	1	1	0	1	0	1	0	1	0
	23	0	6	6	0	6	0	6	6	0	6	0	6	0	6	0
	26	1	0	1	0	0	1	1	0	1	1	0	1	0	0	1
	27	1	0	1	0	0	1	1	0	1	1	0	1	0	1	0
	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	29	1	0	0	1	0	1	1	1	0	0	1	1	0	0	1
	30	1	0	0	1	0	1	1	1	0	0	1	1	0	1	0
	31	5	0	5	0	0	5	5	5	0	0	5	5	0	5	0
	33	1	0	1	0	0	1	1	1	0	1	0	0	1	1	0
	34	1	0	1	0	0	1	1	1	0	1	0	1	0	0	1
	35	6	0	6	0	0	6	6	6	0	6	0	6	0	6	0
	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	37	2	0	0	2	1	1	2	0	2	0	2	2	0	0	2
	38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	42	1	0	1	0	1	0	1	0	1	1	0	1	0	1	0
	43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	44	1	0	0	1	1	0	1	1	0	0	1	1	0	0	1
	45	4	0	0	4	4	0	4	4	0	0	4	4	0	4	0
	46	5	0	5	0	5	0	5	5	0	0	5	5	0	5	0
	47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	48	4	0	4	0	4	0	4	4	0	4	0	0	4	4	0
	49	2	0	0	2	2	0	2	2	0	2	0	2	0	0	2
	50	5	0	5	0	5	0	5	5	0	5	0	5	0	0	5
	51	19	0	19	0	19	0	19	19	0	19	0	19	0	19	0
	52	6	0	0	6	6	0	6	6	0	6	0	6	0	6	0

Figure 5.4: Different types of problem combinations for task type 1 [Complexity = High; Need = Cognitive; Product = Intellectual; Goal = Amorphous].

task_id: 2

		difficult_articulate		irrelevant_results		topknowledge_lack		patient_lack	credibility_uncertain		sources_unaware		toomuch_information		source_unavailable	
		0	1	0	1	0	1	0	0	1	0	1	0	1	0	1
		Count	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count
Problem group	12	0	2	2	0	2	0	2	0	2	0	2	2	0	2	0
	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	17	0	1	1	0	1	0	1	1	0	0	1	1	0	1	0
	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	23	0	3	3	0	3	0	3	3	0	3	0	3	0	3	0
	31	2	0	2	0	0	2	2	2	0	0	2	2	0	2	0
	35	5	0	5	0	0	5	5	5	0	5	0	5	0	5	0
	38	1	0	1	0	1	0	1	0	1	0	1	1	0	0	1
	39	1	0	0	1	1	0	1	0	1	0	1	1	0	1	0
	40	3	0	3	0	3	0	3	0	3	0	3	3	0	3	0
	41	1	0	1	0	1	0	1	0	1	1	0	0	1	1	0
	42	1	0	1	0	1	0	1	0	1	1	0	1	0	1	0
	43	2	0	2	0	2	0	2	2	0	0	2	0	2	2	0
	44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	46	11	0	11	0	11	0	11	11	0	0	11	11	0	11	0
	47	2	0	0	2	2	0	2	2	0	2	0	0	2	2	0
	48	14	0	14	0	14	0	14	14	0	14	0	0	14	14	0
	49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	50	1	0	1	0	1	0	1	1	0	1	0	1	0	0	1
	51	42	0	42	0	42	0	42	42	0	42	0	42	0	42	0
	52	9	0	0	9	9	0	9	9	0	9	0	9	0	9	0

Figure 5.5: Different types of problem combinations for task type 2 [Complexity = Low; Need = Social; Product = Intellectual; Goal = Amorphous].

task_id: 0

		page		people		query		strategy		Other helps	
		0	1	0	1	0	1	0	1	0	1
		Count	Count	Count	Count	Count	Count	Count	Count	Count	Count
Help_group	1	0	1	0	1	0	1	0	1	1	0
	2	0	0	0	0	0	0	0	0	0	0
	3	0	1	0	1	1	0	0	1	1	0
	4	0	0	0	0	0	0	0	0	0	0
	5	0	8	8	0	0	8	0	8	8	0
	6	0	2	2	0	0	2	2	0	2	0
	7	0	1	1	0	1	0	0	1	1	0
	8	0	8	8	0	8	0	8	0	8	0
	9	0	0	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	0	0
	11	3	0	0	3	3	0	0	3	3	0
	12	3	0	0	3	3	0	3	0	3	0
	13	0	0	0	0	0	0	0	0	0	0
	14	6	0	6	0	0	6	6	0	6	0
	15	4	0	4	0	4	0	0	4	4	0
	16	22	0	22	0	22	0	22	0	22	0
	17	1	0	1	0	1	0	1	0	0	1

Figure 5.6: Different types of help combinations for task type 0 [Complexity = Moderate; Need = Cognitive; Product = Intellectual; Goal = Amorphous]..

task_id: 1

		page		people		query		strategy		Other helps	
		0	1	0	1	0	1	0	1	0	1
		Count	Count	Count	Count	Count	Count	Count	Count	Count	Count
Help_group	1	0	14	0	14	0	14	0	14	14	0
	2	0	2	0	2	0	2	2	0	2	0
	3	0	3	0	3	3	0	0	3	3	0
	4	0	1	0	1	1	0	1	0	1	0
	5	0	12	11	1	0	12	0	12	12	0
	6	0	14	14	0	0	14	13	1	14	0
	7	0	2	2	0	2	0	0	2	2	0
	8	0	9	9	0	9	0	9	0	9	0
	9	2	0	0	2	0	2	0	2	2	0
	10	0	0	0	0	0	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	0	0
	12	8	0	0	8	8	0	8	0	8	0
	13	2	0	2	0	0	2	0	2	2	0
	14	7	0	7	0	0	7	7	0	7	0
	15	2	0	2	0	2	0	0	2	2	0
	16	27	0	27	0	27	0	27	0	27	0
	17	6	0	6	0	6	0	6	0	0	6

Figure 5.7: Different types of help combinations for task type 1 [Complexity = High; Need = Cognitive; Product = Intellectual; Goal = Amorphous].

task_id: 2

		page		people		query		strategy		Other helps	
		0	1	0	1	0	1	0	1	0	1
		Count	Count	Count	Count	Count	Count	Count	Count	Count	Count
Help_group	1	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0
	4	0	2	0	2	2	0	2	0	2	0
	5	0	1	1	0	0	1	0	1	1	0
	6	0	4	4	0	0	4	4	0	4	0
	7	0	0	0	0	0	0	0	0	0	0
	8	0	7	7	0	7	0	7	0	7	0
	9	0	0	0	0	0	0	0	0	0	0
	10	1	0	0	1	0	1	1	0	1	0
	11	0	0	0	0	0	0	0	0	0	0
	12	9	0	0	9	9	0	9	0	9	0
	13	1	0	1	0	0	1	0	1	1	0
	14	9	0	9	0	0	9	9	0	9	0
	15	3	0	3	0	3	0	0	3	3	0
	16	43	0	43	0	43	0	43	0	43	0
	17	21	0	21	0	21	0	21	0	0	21

Figure 5.8: Different types of help combinations for task type 2 [Complexity = Low; Need = Social; Product = Intellectual; Goal = Amorphous].

Task 0		Help categories											
		1	3	5	6	7	8	10	12	14	15	16	17
Problem groups	7	N	N	N	N	N	N	N	**Y**	N	N	N	N
	8	N	N	**Y**	N	N	N	N	N	N	N	N	N
	16	N	N	Y	N	N	N	N	N	N	N	N	N
	20	N	N	N	N	N	N	Y	N	N	N	N	N
	23	N	N	N	N	N	N	N	N	N	Y	N	N
	24	N	N	Y	N	N	N	N	N	N	N	N	N
	25	N	N	Y	N	N	N	N	N	N	N	N	N
	28	N	N	Y	N	N	N	N	N	N	N	N	N
	30	Y	N	N	N	N	N	N	Y	N	N	N	N
	31	N	N	Y	N	N	Y	N	N	N	N	Y	Y
	32	N	N	Y	N	N	N	N	N	N	N	N	N
	35	N	Y	N	N	N	Y	N	N	N	Y	Y	N
	36	N	N	N	N	N	N	N	N	N	Y	N	N
	43	N	N	N	N	N	Y	N	N	N	N	N	N
	44	N	N	N	N	N	N	N	N	N	N	N	N
	45	N	N	N	Y	N	N	N	N	N	N	N	N
	46	N	N	N	N	Y	3	N	N	Y	N	Y	N
	48	N	N	N	N	N	N	N	N	Y	N	Y	N
	50	N	N	N	N	N	N	N	N	N	Y	N	N
	51	N	N	N	Y	N	Y	N	N	Y	N	Y	N
	52	N	N	N	N	N	N	Y	N	Y	N	Y	N

Figure 5.9: Problem-Help categories for task type 0 [Complexity = Moderate; Need = Cognitive; Product = Intellectual; Goal = Amorphous].

Task 1		Help categories														
		1	2	3	4	5	6	7	8	9	12	13	14	15	16	17
Problem group	1	Y	N	N	N	Y	N	N	N	N	N	N	N	N	N	Y
	2	N	N	N	N	N	N	N	N	N	N	N	N	N	N	Y
	3	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N
	4	Y	N	N	N	Y	N	N	N	N	Y	N	N	N	N	N
	5	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	N
	6	Y	N	Y	N	Y	N	N	N	N	N	N	N	N	Y	N
	7	Y	Y	N	N	N	Y	N	N	N	N	N	N	N	N	N
	8	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	9	N	N	N	N	N	N	N	N	N	Y	N	N	N	N	Y
	10	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N
	11	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N
	12	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	13	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	14	N	N	N	N	N	N	Y	N	N	N	N	N	N	N	N
	15	N	N	N	N	N	Y	N	N	N	Y	N	N	N	N	N
	16	N	N	N	N	Y	N	N	N	Y	N	N	N	N	N	N
	17	Y	N	N	N	N	N	N	N	N	N	N	Y	N	N	N
	18	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N
	19	N	N	N	Y	N	N	N	N	N	N	N	N	N	N	N
	20	N	Y	N	N	N	Y	N	Y	N	N	N	N	N	N	N
	21	N	N	N	N	N	Y	N	N	N	N	N	N	N	N	N
	22	N	N	N	N	N	N	N	N	N	N	Y	N	N	N	N
	23	N	N	N	N	N	N	N	Y	N	Y	N	Y	N	Y	Y
	24	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	25	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	26	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	N
	27	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	28	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	29	N	N	N	N	Y	N	N	N	N	N	N	N	N	N	N
	30	N	N	Y	N	N	N	N	N	N	N	N	N	N	N	N
	31	Y	N	N	N	N	Y	N	Y	Y	N	N	N	N	N	N
	32	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	33	N	N	N	N	N	N	N	N	N	N	N	N	Y	N	N
	34	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	N
	35	Y	N	N	N	N	Y	N	N	N	Y	N	N	N	Y	N
	36	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	37	Y	N	N	N	Y	N	N	N	N	N	N	N	N	N	N
	38	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	39	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	40	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	41	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	42	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	N
	43	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	44	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	45	N	N	N	N	Y	Y	Y	N	N	N	N	N	N	N	N
	46	N	N	N	N	Y	Y	N	N	N	N	N	Y	N	Y	Y
	47	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	48	N	N	N	N	Y	Y	N	N	N	N	N	N	N	Y	N
	49	N	N	N	N	N	N	N	Y	N	N	Y	N	N	N	N
	50	N	N	N	N	N	Y	N	Y	N	N	N	Y	N	Y	N
	51	Y	N	N	N	N	N	N	Y	N	N	N	N	N	Y	N
	52	N	N	N	N	N	N	N	N	N	N	N	Y	Y	Y	N

Figure 5.10: Problem-Help categories for task type 1 [Complexity = High; Need = Cognitive; Product = Intellectual; Goal = Amorphous].

Task 2

		Help categories										
		4	5	6	8	10	12	13	14	15	16	17
Problem groups	12	N	N	N	Y	N	N	N	N	Y	N	N
	17	N	N	N	N	N	N	N	N	N	N	Y
	18	N	N	N	N	N	N	N	N	N	N	N
	23	N	N	N	Y	N	N	N	Y	N	N	N
	31	N	N	N	N	N	N	N	N	N	N	Y
	35	N	N	N	N	N	Y	N	N	N	Y	Y
	38	N	N	N	N	N	N	N	Y	N	N	N
	39	N	N	N	N	N	N	N	N	Y	N	N
	40	N	N	N	Y	N	N	N	Y	Y	N	N
	41	N	N	Y	N	N	N	N	N	N	N	N
	42	N	N	N	N	N	N	N	Y	N	N	N
	43	N	N	Y	Y	N	N	N	N	N	N	N
	44	N	N	N	N	N	N	N	N	N	N	N
	45	N	N	N	N	N	N	N	N	N	N	N
	46	N	N	N	Y	N	Y	Y	N	N	Y	Y
	47	N	N	Y	N	N	N	N	N	N	Y	N
	48	Y	N	N	N	N	N	N	Y	N	Y	Y
	49	N	N	N	N	N	N	N	N	N	N	N
	50	N	N	N	N	N	N	N	N	N	Y	N
	51	N	Y	Y	Y	Y	Y	N	N	N	Y	Y
	52	N	N	N	N	N	N	N	Y	N	Y	Y

Figure 5.11: Problem-Help categories for task type 2 [Complexity = Low; Need = Social; Product = Intellectual; Goal = Amorphous].

Predicting perceived problem from search behaviors

To predict user-perceived problems from users' search behavioral patterns, we used multiple binomial logistic regressions for each anticipated problem reported by the participants in stages of information seeking. We aggregated search behaviors into three segments: (i) current, (ii) immediate past, and (iii) entire information search segment and associated seeking behaviors. We defined the problem as a dichotomous variable with two options: identified (1) or not identified (0). To investigate the relationship between identified problems and various behavioral features from previous, current, and entire sessions, we used logistic regression with the identified problem as the dependent variable and the behavioral features as independent variables. The tables 5.6- 5.7 list the models' outcomes and goodness of fit. The statistically significant relations ($p < .05$, $p < .01$, $p < .001$) are marked with asterisks on the Table 5.6.

Based on the results presented in Table 5.6, there is a statistically significant association between the difficulty in expressing the current information need and various behavioral measures. These measures include the time spent on SERPs in both previous and current query segments, the number of distinct queries without clicks throughout the sessions, and a negative relationship with the number of bookmarked web pages. This finding suggests that participants' engagement with the visualization of information sources provided by the retrieval systems could lead to a cognitive need to express their thoughts concisely in order to find more helpful information. The amount of time devoted to previous search results may indicate challenges in converting needs into effective queries, encountering unhelpful outcomes, and experiencing difficulties finding valuable sources during subsequent search activities. These observations can be explained by the user's struggle to locate the desired information within the search results. Another possible explanation is that, due to a previous query yielding numerous irrelevant web pages, the user decided to invest time in formulating a query that could retrieve relevant information. Various characteristics of the previous

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

Table 5.6: Coefficients and significance values for each variable in each regression (Significance: *=.05,**=.01, ***=.001) [242]

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

Table 5.7: Logistic Regression: Goodness of Fit (χ^2 scores. Significance: *=.05,**=.01, ***=.001).

task state, including the number of distinct content pages, unique queries, query length, and the number of bookmarks, can predict the subsequent perception of barriers in users' minds, such as perceiving unsatisfactory results, impatience, skepticism towards the displayed information, information overload, and unavailability of the desired information.

Furthermore, significant positive associations exist between the problem of not finding useful results and users' dwell time on SERPs in previous, current, and entire sessions. Additionally, this problem is associated with longer query length in previous and current search segments, retrieving more distinct web pages per query, and a negative association with the number of distinct content pages retrieved throughout the sessions. These results indicate that when participants did not find relevant information in previous search results or perceived a lack of information from the system, they invested extra time and effort in formulating queries to retrieve more helpful information. The results also reveal a negative relationship between impatience and receiving different content pages in the query segment, suggesting that as the search session progresses. Participants do not encounter unique content in previous segments, increasing their impatience in search interaction. Participants who had difficulty expressing a query, encountered many irrelevant results, or needed to be made aware of good information sources tended to spend significantly more time browsing subsequent SERPs. Moreover, if participants distrusted information, they bookmarked significantly fewer pages in the subsequent segment (Table 5.6). Conversely, when participants encountered no problems, they bookmarked more content pages in subsequent query segments.

Surprisingly, at the session level, no browsing behavior feature (prior or current) was significantly associated with a perceived lack of topic knowledge or not knowing all the good or useful sources. These perceptions were more closely related to the prior context of the session. They remained relatively stable throughout the middle of a single search session (Table 5.6). Nonetheless, difficulty in expressing one's information need, encountering many useful or irrelevant results, and being unable to find available information were associated with changes in all three measured components. This suggests that problems are not only influenced by immediate segment strategies (e.g., no clicks on SERPs) but also emerge from users' recognition of general patterns in their search behavior over the session (e.g., few bookmarks) (Table 5.6). Furthermore, the perception and recognition of problems can lead to significant changes in web search behavior, such as increased time spent on future SERPs (Table 5.6). The goodness of fit of the models reported in Table 5.7 indicates the overall statistical significance of the models. It demonstrates the extent to which the predicted and observed frequencies closely match each other, with a better fit indicating a closer match. The table shows that the models predicting perceived problems related to query articulation, useless and irrelevant information, and unavailability of information can significantly outperform random chance based on users' previous, current, and session behaviors. Moreover, behaviors in earlier search sessions and throughout the entire session are better predictors of the lack of patience and the problem of excessive information (Table 5.6).

Observing current search behaviors makes it possible to infer the presence of these problems, such as difficulty in expressing needs through search queries, encountering unsatisfactory results, doubt regarding the trustworthiness of the information displayed, limited awareness of reliable sources, and unavailability of desired information. The study also suggests that these perceived problems arise not only from immediate search strategies (e.g., lack of clicks on search engine result pages) but also from users' overall perception and recognition of patterns in their search behavior throughout the session (e.g., limited use of bookmarks). Such perceived problems in search can lead to notable changes in user behavior during subsequent query segments, including increased time spent on future search engine result pages.

Predicting helps from behaviors and problems

The perceived problems of a searcher are closely related to the assistance they believe is most suitable for addressing their information needs [242]. We conducted a classification to investigate the connection between a person's perceived problem and their perceived help. We annotated query segments with problems and corresponding help. Our analysis focused on determining which aspects of a searcher's session contributed most to accurately classifying their perceived best help at any given moment. We employed a simple logistic regression classifier, splitting the data into 80% for training and 20% for testing. We compared its performance to three simple baselines: a random baseline, a most frequent labeling baseline (constantly labeling a problem as absent), and a baseline that always labels a problem as present. Tables 5.8-5.11 provide an overview of the results.

In general, the characteristics of a search influence the perceived assistance at a given moment. However, accuracy scores might suggest otherwise. Naive baselines, such as the most frequent labeler, often exhibit comparable or significantly higher accuracy, indicating that the selected behaviors may not predict perceived assistance. This phenomenon arises due to the imbalanced nature of the problem, as a single perceived help occurrence is not dominant during sessions. Precision scores, on the other hand, indicate the ability of a classifier to identify when a person requires assistance correctly. The number of problems and the combined browsing features can determine if no assistance is needed. In contrast,

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

Table 5.8: Accuracy scores for predicting the currently reported help using previous query segment features, session features, the currently reported problem, and all features combined.

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

Table 5.9: Precision scores for predicting the currently reported help using previous query segment features, session features, the currently reported problem, and all features combined.

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

Table 5.10: Recall scores for predicting the currently reported help using previous query segment features, session features, the currently reported problem, and all features combined.

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

Table 5.11: F1 scores for predicting the currently reported help using previous query segment features, session features, the currently reported problem, and all features combined.

session features can identify when a page provides the desired help. Moreover, the previous segment or the session itself can indicate if a person is the desired help, and all the features combined can determine if a query is the desired help. Evaluating the entire prior session can ascertain whether a strategy is the desired help and dissatisfaction with the search system can be inferred from the entire session. The F1 scores provide further validation of these precision scores.

5.4 Discussion and Implications

This study investigates the interconnection between users' implicit and explicit search behaviors, perceived constraints and barriers encountered during online information search, and preferred methods of assistance to overcome these barriers at various stages of the search process. The data analysis reveals several interesting observations, some of which align with existing studies and common knowledge. In contrast, others provide nuanced insights into the barriers individuals face during different search stages, their recognition of information needs, preferred forms of help, and search behaviors.

The findings of this study demonstrate statistically significant associations between individuals' perceived problems and their preferred forms of assistance during the search process, and these relationships can be elucidated through users' search behaviors. Thus, predicting users' perceived problems at different stages of the information search process becomes possible by mapping their implicit and explicit search behaviors.

Moreover, the study aims to develop a comprehensive and quantifiable understanding of users' cognitive focus state at any given moment by conceptual mapping and empirically measuring these three aspects within the information search process. The results indicate that combining information on perceived problems requested assistance, and search behaviors at a specific moment; it is feasible to infer users' momentary cognitive focus of attention. For instance, if a user realizes a lack of trust in the retrieved information and requests more useful pages while simultaneously decreasing the number of bookmarked pages, these actions could indicate a need for more relevant information. Additionally, this study provides insights into the relationships among task type, problem type, and preferred forms of assistance (Figures 5.9, 5.10, 5.11), which will be valuable for further conceptualizing and formalizing the task representation model.

Chapter Summary

- The study explores the connection between users' implicit and explicit searching behaviors, perceived constraints and barriers during online information search, and preferred help to overcome those barriers.

- Statistically significant associations are found between users' perceived problems and preferred helps, which can be explained through search behaviors.

- Different clusters of search behaviors can indicate struggling users at various stages of the search process.

- Features of the previous task state, such as distinct content pages, unique queries, query length, and number of bookmarks, can predict users' perception of barriers in their search process.

- Users' perceptions of problems can arise from both immediate search strategies and general patterns of their search behavior throughout the session, leading to significant behavioral changes in future query segments.

- Desired help at a given task state is not solely determined by perceived problems but is influenced by the strategies employed over the entire search session.

- Combining perceived problems, help, and search behaviors allows for inferring users' momentary cognitive focus of attention during the search process.

Chapter 6

RESEARCH QUESTION 3: STUDY 3: CHARACTERIZE TO CONSTRUCT UNIFIED TASK REPRESENTATIONS

A complex search task is an interactive process that often consists of multiple search sequences or stages of users' iterative interactions with a search system and the retrieved information objects (e.g., defining the problem, selecting the source, executing a query, examining retrieving results [189]). Also, it is a series of transitions from one task state to another state as represented by searchers' iterative explorations to find solutions for the problematic situations behind the task [212].

By utilizing knowledge gained from Chapter 4 and Chapter 5, we can conceptualize a holistic picture of complex search tasks as reflections of patterns of task intentions [171], behavioral actions in search session segments [108], task strategies, problems and helps and other aspects of search interactions. To operationalize, we then build a context-agnostic representation of the task (at a granular level) that can be transferable to any search context and any level – macro to micro (such as task levels described in [258]) in Chapter 7.

6.1 A Holistic Notion of Task: Conceptualize Users' Cognitive Focus or Task State

The user's cognitive focus during the information search process is a dynamic state of mind that evolves over time and can be observed through their observable physical and cognitive activities, such as the problems they encounter and the assistance they require to understand and address those problems [173, 242]. This cognitive focus represents the perceived gaps in the user's knowledge and understanding, which continuously change as they progress through the information seeking process.

When users face a problematic situation without sufficient knowledge to resolve it, their motivation behind intention or information need may be ill-defined, hindering their progress in the search process and requiring additional support. As users advance through the process, their focus of attention evolves, and they develop a clearer understanding of their needs, leading to changes in their judgments of information based on their increasing knowledge and comprehension of the problem [239, 147, 145, 305].

Variations in humans' cognitive focus states, influenced by socio-cultural contexts and situations, result in different motivations for information needs, strategies, local task goals, and challenges. To map users' cognitive focus of attention or task states at different stages of the information search process, we draw upon theoretical and methodological works, including Dervin's sense-making approach, which emphasizes the situation-gap-use framework [83, 84, 225, 128]. This mapping identifies various situations that create information needs or "gaps" [128] based on users' perceived problems or barriers in expressing their needs and seeking information, as well as the preferred assistance they seek to address the problem. It also considers users' actions, such as formulating queries and selecting retrieved documents. When faced with a gap in the information search process, users engage in reflective thinking about potential problems and strategies to bridge the gap.

Although some argue that the cognitive focus of attention cannot be concretely measured or directly observed due to its subjective nature, this dissertation asserts that it is a task state that is consciously experienced and immediately recognizable, making it possible to observe and measure through users' expressible physical and cognitive activities, including their tactics, strategies, actions, problems, and the assistance they seek while seeking information. Searchers' cognitive focus states are often implicit in their interactions with systems, as evidenced by their queries, browsing behavior, dwell time, clicks, and mouse movements [242, 173, 2, 172, 52, 73]. Therefore, analyzing search behavioral patterns can provide insights into the latent cognitive processes underlying users' information seeking activities. For example, in an exploratory search session (presented in Figure 6.1), a searcher wanted to "learn about Swahili dishes and how to cook them" [57]. In the first query iteration, the searcher's expressed goal was broad, and they clicked several results. However, after gaining some knowledge of the task during the first and second query iterations and checking several result pages, the searcher's visceral state of information need became more specific, which was prominent in the third query. During the search episode, we can observe that the searcher's intent shifted from "swahili dish" to "swahili food" and ended with " "Swahili recipe." In other words, the searcher's cognitive state shifts from exploring and exploiting the problem space.

6.2 Task States

We have introduced a concept called the "cognitive focus" or "task state" to gain a comprehensive understanding of task [243]. This concept encompasses both the cognitive and behavioral aspects of complex tasks, enabling us to model the transitions between different task states as a complex search process. At a detailed level, a complex search task can be seen as a sequence of momentary cognitive focuses or task states. Each cognitive focus state is characterized by the user's local task motivation, the encountered problem, and the preferred help [243]. Users also employ a set of tactics or strategies to accomplish the task [108]. For instance, a task state could be represented as "SA-IO-SR" (Seeking advice - Information overload - Source recommendation), which can be reflected in the actions taken by users [243]. Our model builds upon Liu's previous work [170], presenting an improvement over their model. Figure 6.2 visually represents users' cognitive focus or task state. For example, based on the task-problem-help mapping described in Chapter 5, a task state can be a combination of the motivation to find specific information, encountering the problem of information overload, preferring source recommendation for help, and performing actions such as clicking on documents.

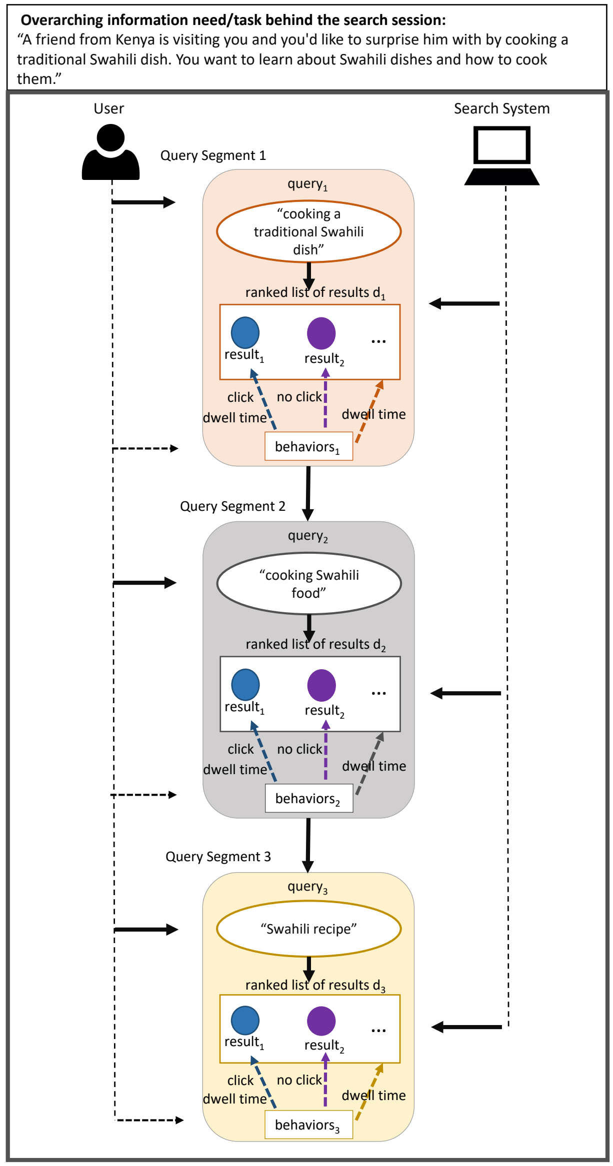

Figure 6.1: An example search session which shows multiple query segments from TREC Session Track 2014.

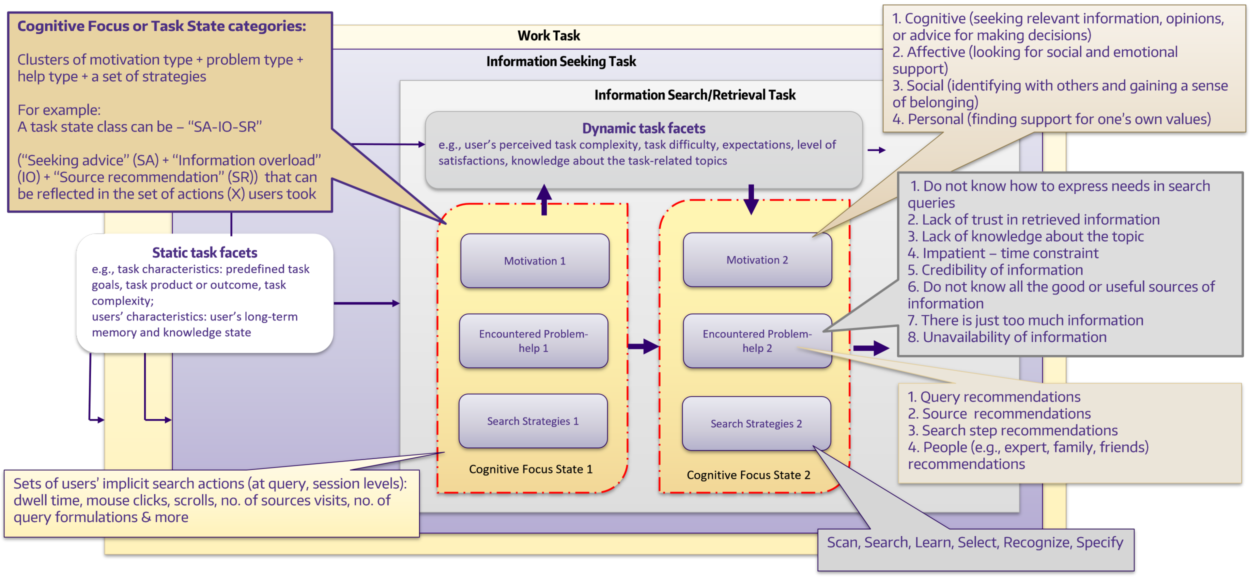

Figure 6.2: A Holistic Task Model based on users' motivation state [171], problem-help state [242] and strategies [108] [108].

6.3 *Preparing Ground-Truth Data: Query Log Annotation Study: Identify Task States*

To determine task states, we employed the search query log collected from the lab study conducted with 37 participants in Study 2 (for detailed information, refer to Chapter 5) and the task-problem-help mapping [173]. To ensure accuracy, we randomly selected 10% of search sessions from each task type and assigned two assessors to annotate the task state for each query segment independently. These annotations were based on the task states defined using the above-mentioned formula. The process consisted of two phases, and after each phase, the assessors resolved any discrepancies through discussions. We obtained 768 task states with these assessments by considering all aspects and distinct options.

For the remaining dataset, we employed the *PAM* clustering analysis technique on the lab study data to extract potential task states [133]. *PAM* is an unsupervised learning method that identifies a sequence of centrally located medoids within data clusters. It achieves this

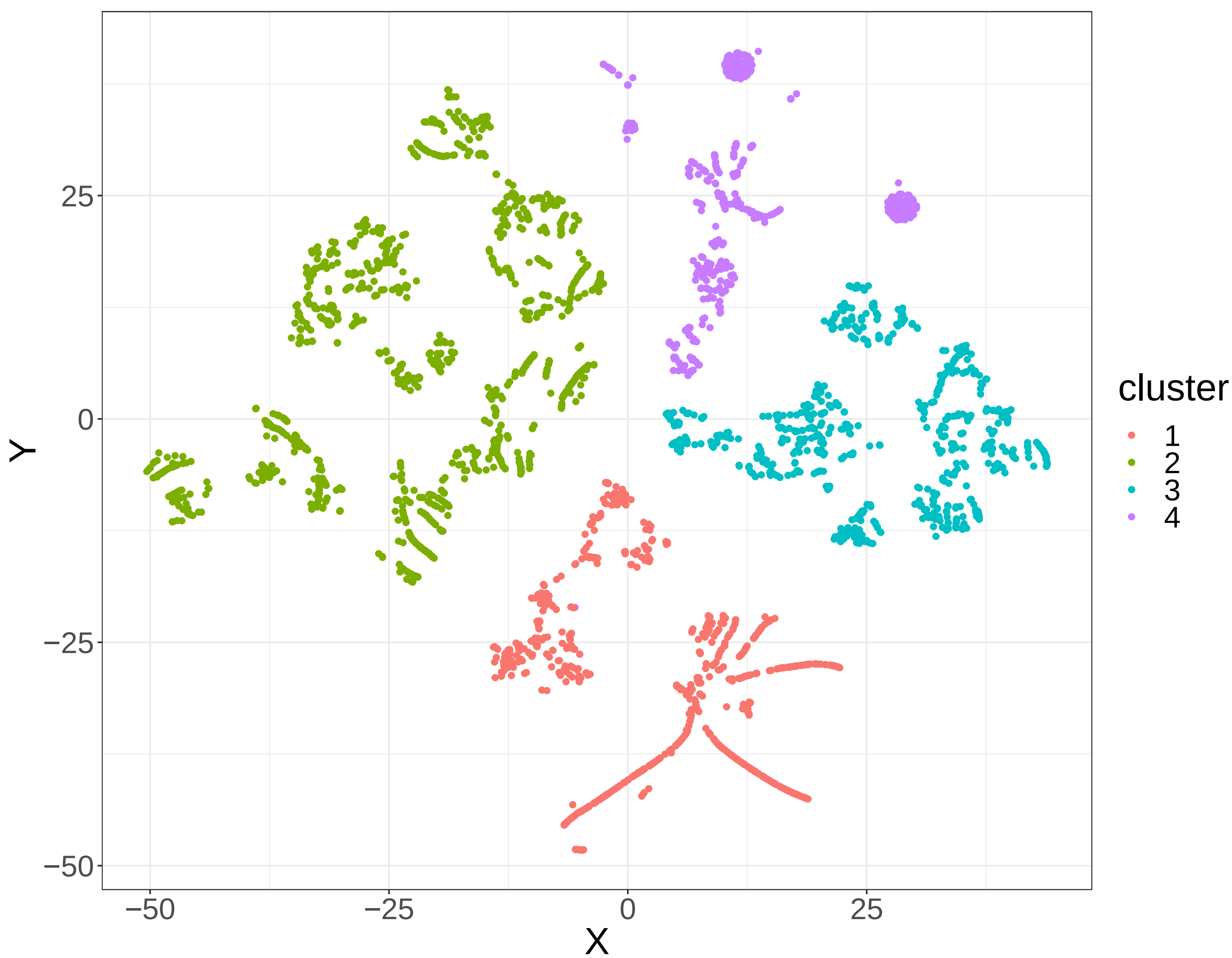

Figure 6.3: Visualization of the four from multiple task state clusters, predicted via PAM analysis of the behavioral features. We used t-Distributed stochastic neighbor embedding (t-SNE) to project the behavioral features onto a 2-dimensional space.

by minimizing the average dissimilarity between each data point and its closest selected data point. The optimal number of task clusters was determined using silhouette analysis. Figure 6.3 illustrates four task state clusters obtained from the behavioral features in a low-dimensional space. While imperfect, the clusters exhibit visible separation, confirming their relevance.

In the clustering analysis, information seeking motivations, in-situ problems, and needed help types were considered separate elements within the vectors representing unique task states. We validated the task state clusters by comparing them with output from two independent human assessors who manually annotated the state labels. After that, To measure the validity of task state labels, we computed three *Cohen's Kappa coefficients*, between (i) the two annotators, (ii) the annotator A and the clustering algorithm, and (iii) the annotator B and the clustering algorithm. Cohen's Kappa agreement between all three pairs of algorithmic and human annotations was above 70%. 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.

6.3.1 Predicting Task States from User Behaviors

Our study used behavioral signals to conduct a small predictive analysis of task states. We developed multiple classifiers based on different user behavioral signals to achieve this. These classifiers were trained and evaluated using an 80/20 split on training and testing data. We compared their performance against two baseline models: 1) a random baseline and 2) a most frequent labeling baseline. The accuracy scores for the prediction analysis can be found in Table 6.1.

The behavioral features considered in our analysis include various aspects of search behavior. These features encompass the following: *query behavior*, such as query length and query reformulation type; *browsing behavior*, including the number of clicks and the number of content pages visited; *dwell time (second)*, which consists of the mean dwell time on each SERP, the mean dwell time on each content page, and the total dwell time on content pages; and *usefulness judgment*, represented by the number of bookmarks. Using these fundamental measures, we constructed three types of feature sets for predicting task states: 1) behavioral measures within the *current query segment* associated with the target task state, 2) *session-level behavioral measures* prior to the current query segment, and 3) a *combination* of the first and second sets. The session-level behavioral measures were computed as the average values of the behavioral measures for the corresponding session before the current query segment.

The results presented in Table 6.1 demonstrate that the top-performing classifiers significantly outperform the corresponding baseline models regarding overall accuracy in predicting task states. Constructing classifiers solely using behavioral data from the current query segment consistently yields better results than the baseline models. However, the best outcomes are achieved by combining the session-level data from previous query segments with the data from the current segment. These findings highlight the predictive potential of search behavioral signals in capturing dynamic task states, which subsequently led to the development of our unified task representation model discussed in Chapter 7.

Chapter Summary

- Our task state conceptualization covers both the active, motivational, and unanticipated situational dimensions of task states.
- We analyzed task states and observable behaviors by using clustering algorithms to extract task states from annotation data. We then verified the state labels through manual annotations and assessments.
- Intelligent search systems can detect task states in real-time and use this information to assist users in completing tasks more effectively.

Table 6.1: Accuracy score of task state prediction. *Note*: Significant values indicate whether the predictor is significantly better than the best baseline (*:p<.05, **:p<.01). The best performers are boldfaced.

Classifier	**Current seg.**	**Prev. session**	**All data**
Support Vector Machine	0.467	0.485	0.483
Random Forest	0.689**	**0.693****	**0.692****
K-nearest Neighbors	0.634**	0.620**	0.611**
Most Frequent	0.454	0.454	0.454
Random	0.442	0.442	0.442

Chapter 7

RESEARCH QUESTION: STUDY 3: A UNIFIED TASK REPRESENTATION

In this chapter, we tackle the fundamental issue of task representation by proposing a graph-based approach to model task embedding. Our method considers users' task states during the search process and is applicable to a wide range of search and retrieval scenarios, regardless of the search modes and modalities employed. A key innovation of our approach is the task-context representation learning, which allows the system to uncover hidden task information from constrained search behaviors.

7.1 Introduction

Drawing on existing research, we propose an alternative approach to representing tasks that captures the relationships between diverse entities, including user actions and information artifacts (such as queries and documents). One viable solution for achieving this lies in using a graph-based structure. However, the challenge lies in transforming a conceptual representation into a practical implementation. In light of this, we introduce a task embedding technique called Heterogeneous Graph Neural Network (HGNN), which enables an understanding of task states by capturing users' search processes and evolving search situations. This involves encoding users' query segments into an HGNN-based task embedding, incorporating semantic, temporal, and other behavioral information derived from user and system interactions. We capture the users' task states by measuring the networked relation between the vector representations of users' search activities, queries, and documents. Our main contribution lies in operationalizing the task model, which encodes the structure of search episodes and relevant contextual information into low-dimensional feature vectors. These vectors can then be leveraged by statistical modeling techniques, enabling the implementation of effective training and modeling approaches for search and retrieval systems. To evaluate the quality of the extracted task representations and the effectiveness of our task extraction methodology, we conduct experimental, quantitative evaluations, as described in Chapters 9 and 10.

7.2 Problem Formulation and Task Embedding Model

According to [209], empirical research indicates that knowledge graphs and graphical modeling are particularly effective in capturing the characteristics of search tasks. Therefore, in this study, we visualize the interaction between users and the system by representing a user-system interaction log as a complex network structure. This network includes information on users, systems' actions (such as query submission and clicks on documents), and reactions (such as analysis, retrieval, and display of ranked information objects related to users' queries). In this network, users, queries, and documents are represented as vertices, and clicks and dwell time serve as edges, similar to a social network. To create this network graph representation of the search session, we specifically employ two graph neural network models proposed by [272] and [96] [1].

Our model assumes that the progression of task states in query sequences depends on the semantic information of the queries, search sessions, and corresponding search factors, such as searchers' interactions with the systems (e.g., search behavior patterns like the number of queries, number of clicked results, click number per query, and time duration of query sequences), as well as other session-specific features. In our approach, we first develop an embedding model representing the search sessions as networked graphs. This graph-based embedding module is described in Part I of our methodology. Part II introduces an automatic

[1] It should be noted that we have incorporated relevant aspects of these two models into our study. We recommend referring to the original papers for more detailed descriptions.

state detection module that categorizes training query segments into task state categories based on their semantic and behavioral feature patterns. We evaluate our proposed method on two search session datasets and compare its performance with two alternative task state detection models. The experimental results demonstrate that our method outperforms the alternatives by accurately identifying task states embedded in query logs and detecting only task states.

7.2.1 Part I: HGE-GNN based Search Session and Query Segments Embeddings

In this section, we provide formal definitions of key terminologies pertinent to our application of metapath-driven heterogeneous graph neural networks [272, 96]. Subsequently, we present a comprehensive, step-by-step outline of the entire process.

Preliminary

Definition 1: Heterogeneous Graph

A heterogeneous graph of search session can be represented as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is the set of node objects with a node type mapping function $\phi : \mathcal{V} \rightarrow \mathcal{A}$, and $\mathcal{E}$ is a set of edges connecting the nodes in V: $\mathcal{E} \subseteq V \times \mathcal{V}$ with an edge type mapping function $\psi : \mathcal{E} \rightarrow \mathcal{R}$. $\mathcal{A}$ and $\mathcal{R}$ denotes the predefined sets of node types and edge types, respectively [230].

Definition 2. Metapath

Each node type is connected by one or multiple methapath instances. A metapath p is defined as an ordered sequence $\mathcal{A}_1 \xrightarrow{\mathcal{R}_1} \mathcal{A}_2 \xrightarrow{\mathcal{R}_2} ... \xrightarrow{\mathcal{R}_l} \mathcal{A}_{l+1}$, which describes a composite relation $R = R_1 \circ R_2 \circ ...R_l$ between node types $\mathcal{A}_1$ and $\mathcal{A}_{l+1}$ [96].

Definition 3. Metapath-based Neighbor

Given a set of metapath $\mathcal{P}$ of a heterogeneous graph, the metapath-based neighbors $\mathcal{N}_v^p$ of a node v is defined as the set of nodes that connect with node v via metapath instances of $\mathcal{P}$.

A neighbor node connected by two different metapath instances is regarded as two different nodes in $\mathcal{N}_v^p$.

Definition 4. Node Attributes

Each node type v has a number of attribute features $\{\mathbf{x}_v, \forall v \epsilon \mathcal{V}\}$ including textual content (e.g., query text, item URL) and other search-system interaction signals (e.g., number of clicks or dwell time) recorded in the search log.

Definition 5. Heterogeneous Graph Embedding

Given a heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with node attribute matrices $\mathcal{X}_{A_i} \epsilon R^{|\mathcal{V}_{A_i}| \times d_{A_i}}$ for node types $A_i \epsilon \mathcal{A}$, heterogeneous graph embedding is the task to learn the d-dimensional node representations $\mathbf{h}_v \epsilon R^d$ for all $v \epsilon \mathcal{V}$ that are able to capture rich structural and semantic information involved in $\mathcal{G}$ [96].

For instance, let's consider the heterogeneous search session graph depicted in Figure 7.1. This graph encompasses three distinct types of nodes: task or sub-task topics, query segments consisting of query-item pairs, and clicked items. By leveraging this graph, we can establish connections between these nodes using a minimum of three metapaths. These metapaths capture composite relationships among the topics, queries, and items. As an illustration, Item 1 serves as a metapath-based neighbor of Item 3, with their connection being established through the metapath instance I-T-Q-I.

Model Architecture

To construct query segment embeddings for our task state prediction objective, we utilize the metapath aggregated graph neural network framework for heterogeneous graph embedding introduced by Fu et al. [96]. Following their approach, the graph construction comprises three key components: node content transformation, intra-metapath aggregation, and inter-metapath aggregation [96]. Figure 7.2 provides a simplified graphical representation of the

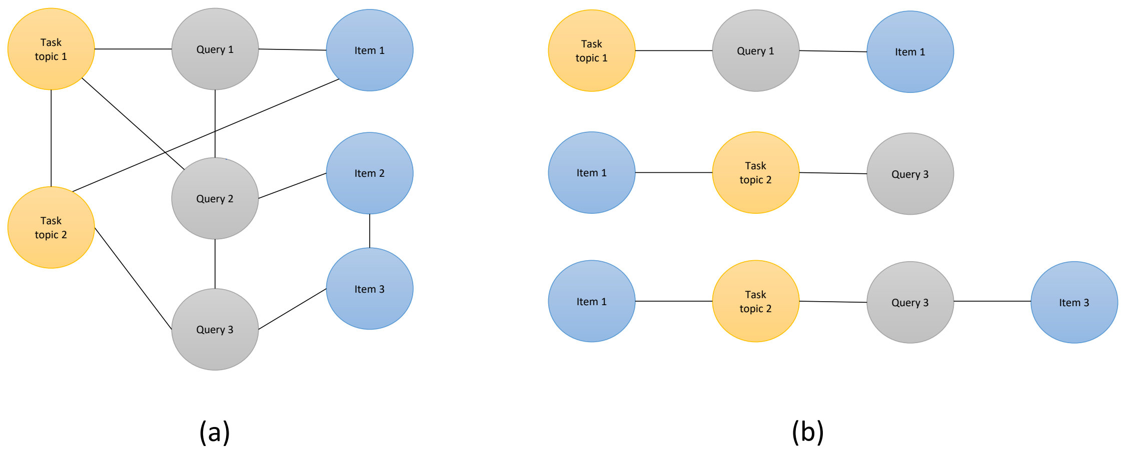

Figure 7.1: An example of heterogeneous search session graph with three types of nodes (i.e., task topics, query, and clicked items) (a), and three different metapath instances (i.e., Topic-Query-Item (T-Q-I), Item-Topic-Query (I-T-Q), Item-Topic-Query-Item (I-T-Q-I)) (b).

embedding process for an individual node.

Node Content and Attribute Transformation

Our initial objective is to represent node features of different types in a unified latent vector space and learn d-dimensional node representations, denoted as $\mathbf{h}_v \in R^d$, for all nodes $v \in V$. This allows us to capture both the structural and semantic information present in a search session graph G [96]. To achieve this, we follow the approach presented in [272] and [96], where our model applies node type-specific transformation techniques to embed the textual content and other attributes of each node into a shared low-dimensional vector space. For a node $v \in \mathcal{V}_A$ of type $A \in \mathcal{A}$, we have

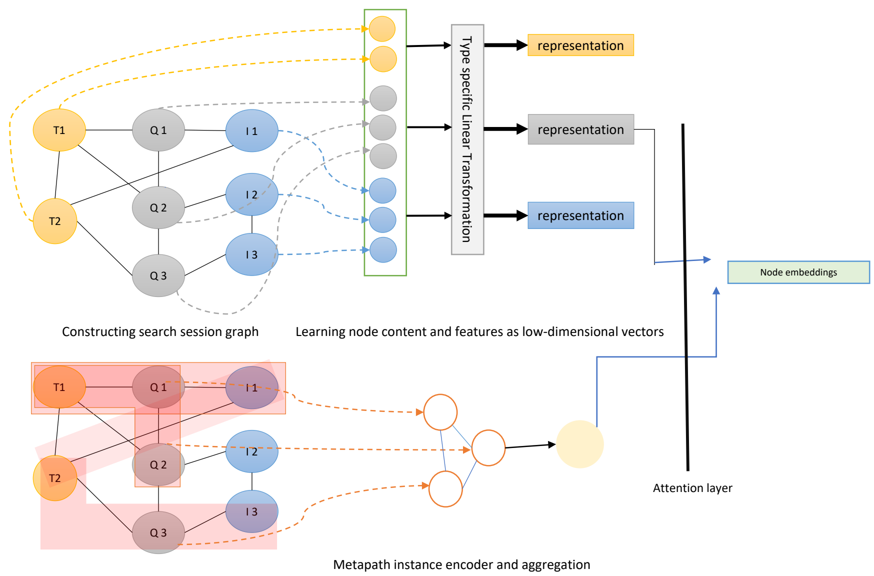

Figure 7.2: A simplified example of step-by-step embedding for a single search session node.

$$\mathbf{h}'_v \leftarrow \mathbf{W}_A \cdot \mathbf{x}_v^A, \forall v \epsilon \mathcal{V}_A \tag{7.1}$$

where $\mathbf{x}v \in R^{dA}$ represents the original feature vector of node v, and $\mathbf{h}'_v \in R^{d'}$ denotes the projected latent vector for node v. Additionally, $\mathbf{W}_A \in R^{d' \times d_A}$ corresponds to the weight matrix specific to nodes of type A. In our approach, we employ two types of linear transformations for each node type v: attribute-specific embedding and text-based embedding. These transformations facilitate the generation of rich contextual embeddings. Specifically, the text-based embedding captures the underlying semantics of the text content associated with a node. Following this operation, the projected features of all nodes share the same dimension.

Intra-metapath Aggregation

In order to capture the structural embedding of search sessions, such as the similarity between a node's query and its neighboring query or item nodes, we employ the metapath aggregation approach proposed by Fu et al. [96]. This approach allows us to quantify the direct impact of one node type on other nodes within the graph. In this layer, using a given metapath set $\mathcal{P}$, we calculate the structural and semantic information embedded in metapath instances $p(v,u)$ that connect a target node v, a metapath-based neighbor $u \in N_v^p$, and the contextual information in between. To encode each metapath instance through intra-metapath aggregation, we employ a linear $MEAN$ encoder function. This function computes the element-wise mean of all the node vectors along the metapath instance $p(v,u)$ and appends it into a single vector after applying a linear transformation.

$$\mathbf{h}_{p(v,u)} = \mathbf{W}_p \cdot MEAN\left(\{\mathbf{h}'_t, \forall t \epsilon p(v,u)\}\right) \tag{7.2}$$

Next, we combine and embed all the extracted metapath instances into low-dimensional vector representations for a given node v. As different metapath instances can have varying degrees of influence on the representation of the target node, we calculate weights for each metapath instance p related to the target node v using a graph attention layer [279]. To achieve this, we learn a normalized importance weight $\alpha^{\mathcal{P}}vu$ for each metapath instance $p(v,u)$ connecting node v to node u, and then perform a weighted sum of all the instances. Finally, the representations of the metapath instances on node v are passed through an activation function $\sigma(\cdot)$. We employ K independent attention mechanisms and concatenate their outputs, where $[\alpha^{\mathcal{P}}vu]_k$ represents the normalized importance of metapath instance $p(v,u)$ to node v in the k-th attention head. In this step, given the projected feature vectors $\mathbf{h}'_u \in R^{d'}$ for all nodes $u \in \mathcal{V}$ and the set of metapaths $\mathcal{P}_A = P_1, P_2, ..., P_M$ starting or ending with node type $A \in \mathcal{A}$, the intra-metapath aggregation generates M metapath-specific vector representations of the target node $v \in \mathcal{V}_A$ [96].

Inter-metapath Aggregation

Following the aggregation of nodes and edges within each metapath, we proceed to combine the semantic information revealed by all metapaths using an inter-metapath aggregation layer. For a node type A, we obtain $|\mathcal{V}_A|$ sets of latent vectors: $\mathbf{h}_v^{P_1}, \mathbf{h}_v^{P_2}, ..., \mathbf{h}_v^{P_M}$ for $v \in \mathcal{V}_A$, where M represents the number of metapaths for type A. To perform inter-metapath aggregation, we compute the element-wise mean of these node vectors while leveraging an attention mechanism to assign different weights to different metapaths.

$$\beta_{P_i} = \frac{exp\left(\mathbf{q}_A^T \cdot \left(\frac{1}{|\mathcal{V}_A|}\sum_{v\epsilon\mathcal{V}_A} tanh\left(\mathbf{M}_A \cdot \mathbf{h}_v^{P_i} + \mathbf{b}_A\right)\right)\right)}{\sum_{P\epsilon\mathcal{P}_A} exp(e_P)} \tag{7.3}$$

$$[\mathbf{h}_v^{\mathcal{P}_A}]^l \leftarrow \sum_{P\epsilon\mathcal{P}_A} \beta_P \cdot [\mathbf{h}_v^P]^l, \forall v\epsilon\mathcal{V}_A \tag{7.4}$$

where $\mathbf{M}_A\epsilon R^{d_m \times d'}$ and $\mathbf{b}_A\epsilon R^{d_m}$ are learnable parameters, and where $\mathbf{q}_A\epsilon R^{d_m}$ is the parameterized attention vector for node type A. β_{P_i} is the relative importance of metapath P_i to type A's nodes. Finally, the model employs an additional linear transformation with a nonlinear function to project the node embeddings to the vector space with the desired output dimension:

$$\mathbf{h}_v^i = \sigma\left(\mathbf{W}_o^l \cdot \left[\mathbf{h}_v^{P_A}\right]^l\right), \forall v\epsilon\mathcal{V}_A, \forall A\epsilon\mathcal{A} \tag{7.5}$$

where $\sigma(\cdot)$ is an activation function, and $\mathbf{W}_o\epsilon R^{d_o \times d'}$ is a weight matrix [96]. Thus, we concatenate both content and metapath aggregated embeddings together to get the final vector representation of a node. By following this approach, our model can effectively capture the structural and textual content information from neighboring nodes as well as the metapath context in between. This allows us to learn and concatenate diverse low-dimensional embeddings, ultimately generating a comprehensive representation vector for each node $v \in V$. The overall forward propagation process is depicted in Algorithm 1.

Algorithm 1: Steps in our Model [96]

Input: Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$

Node types $\mathcal{A} = \{A_1, A_2, ..., A_{|\mathcal{A}|}\}$

Metapaths $\mathcal{P} = \{p_1, p_2, ..., p_{|\mathcal{P}|}\}$

Node features $\{\mathbf{x}_v, \forall v \epsilon \mathcal{V}\}$

The number of attention heads K

The number of layers L

Result: The node embedding $\{\mathbf{z}_v, \forall v \epsilon \mathcal{V}\}$

initialization;

for *node type* $A \epsilon \mathcal{A}$ **do**

  node content transformation (7.1) ;

end

for $l = 1...L$ **do**

  for *node type* $A \epsilon \mathcal{A}$ **do**

    for *metapath* $P \epsilon \mathcal{P}_A$ **do**

      for $v \epsilon \mathcal{V}_A$ **do**

        Calculate $\mathbf{h}^l_{P(v,u)}$ for all $u \epsilon \mathcal{N}^P_v$ using the metapath instance encoder function (7.2);

        Combine extracted metapath instances (7.3) ;

      end

    end

    Calculate the weight β_P (7.4);

    Fuse the embeddings from different metapaths (7.5) ;

  end

  Layer output projection ;

end

$\mathbf{z}_v \leftarrow \mathbf{h}^L_v, \forall v \epsilon \mathcal{V}$

7.2.2 Part II: Task State Identification

Once the final node representations are generated, we can utilize them as input features for inferential or predictive models to infer task states. The underlying assumption is that each node and edge type corresponds to a specific task state. This allows us to extract task states based on similar patterns observed in node representations. Since task state detection models heavily depend on labeled data, we employ a combination of supervised and unsupervised learning techniques to evaluate our graph-based embedding features.

We optimize the model weights for supervised learning by minimizing the cross-entropy loss through back-propagation and gradient descent. This process enables us to learn meaningful node embeddings for heterogeneous graphs [96]. The cross-entropy loss function for semi-supervised learning is formulated as follows:

$$\mathcal{L} = -\sum_{v \epsilon \mathcal{V}_{\mathcal{L}}} \sum_{c=1}^{C} \mathbf{y}_v[c] \cdot log \mathbf{h}_v[c] \tag{7.6}$$

where $\mathcal{V}_{\mathcal{L}}$ is the set of nodes that have labels, C is the number of classes.

7.3 Experiments

In this section, we present experiments to illustrate the efficacy of our proposed search session embedding. The following research questions guide the experiments:

> **RQ3** What approaches can be employed to effectively represent and express tasks in a manner that is search context agnostic?
>
> **RQ3.1** How do the graph-based embedding features perform in identifying plausible task states?
>
> **RQ3.2** How do we understand the task state representation capability of our HGE-GNN-based approach?

7.3.1 Datasets

Apart from the ground-truth Study 2 search log data (Chapter 6) and the simulated search session data (Chapter 8) both already labeled with task states, we also include two benchmark IIR datasets to evaluate the performance of our proposed approach and compare it with alternate task state detection models. The first dataset is the TREC-Session Track 2014 dataset [57], while the second dataset is a search session dataset collected by Liu et al. [174], referred to as the KDD'19 dataset. Table 7.1 provides a summary of the two datasets.

The TREC 2014 Session Track dataset consists of 1,257 search sessions, with each session containing 3,276 query segments and 10 relevant documents for each query. On the other hand, the KDD User Study dataset comprises 450 search sessions and 1,548 query segments, which are associated with 9 exploratory search tasks. From both datasets, we extracted various search behavioral features at the session level as well as the query segment level. These features include browsing behavior metrics (e.g., number of clicks, number of visited content pages, dwell time), query-related features (e.g., query length, query reformulation type, count of query terms, query similarity, presence of new and unique terms, query order), and item-related features (e.g., number of items retrieved for each query, terms extracted from item titles, snippets, and URLs, number of clicks on items). Additionally, we incorporated other contextual features derived from the data. The textual content associated with queries and documents was represented as bag-of-words representations. Table 7.2 provides an overview of some attribute features (behavioral measures in the current query segment as well as throughout the session) that were extracted and utilized in the task state modeling. Following the same procedure mentioned in Chapter 6, we added task labels to these datasets.

Table 7.1: Summary of the datasets including nodes, edges, and some example metapath instances defined in experiments

Dataset	Node	Edge	Metapath
TREC'14	#user sessions 1,257 #query segments 3276 #topics 62 #items 32,760	Searcher searchers for a topic, Searcher submits a query, Searcher clicking an item, searcher clicking an item under a query, searcher clicking an item, under a query and topic, Query belongs to a topic, Item belongs to a topic	Topic-Query-Item, Searcher-Query-Item, Query-Query, Query-Item, Query-Topic-Query, Item-Topic-Item
KDD'19	#user sessions 450 #tasks 9 #query segments 1,548		

Table 7.2: Some of the features considered for the task state detection.

Notation	**Description**
query attributes	
query terms	# unique terms in a query
query similarity	similarity between current query and the previous query
query order	the order of the current query within the search session
query reformulation type	type of reformulation (e.g., new, partial matching)
query length	# terms used in an issued query
query text	content of the query
query tf-idf	query text representation
query topic	topic of the query
query-item attributes	
item text	title of the item
item tf-idf	item representation
number of clicks	number of clicks in query segment
number of items clicked	number of items clicked
dwell time (second)	total dwell time within the query segment
mean dwell time per result list	average dwell time on result list
mean dwell time per item	average dwell time on item
total dwell time on items	total dwell time on item
action count	total number of actions in query segment

7.3.2 RQ3.1: Detecting Task States

To address RQ3.1, we tackle the task of predicting task states using both supervised and unsupervised approaches.

Generating Graph-based Input Features

We created query segment embeddings for each dataset by leveraging the interactions between various node types. The nodes, edges, and example metapath instances used in the experiments are summarized in Table 7.1. Additionally, Table 7.2 showcases some of the search behavioral features employed in the HGN-GNN model during the node transformation stage to generate contextual representations of query segments. It is worth noting that the availability of these features depends on the specific search scenarios and stages of search sessions.

For training our graph embedding model, we followed the methodology described in [96]. We set the dropout rate to 0.5 and utilized the Adam optimizer with a learning rate of 0.005 and a weight decay of 0.001. The model was trained for 100 epochs, and early stopping with a patience of 30 was applied. Moreover, we configured the model with 8 attention heads, a dimension of 128 for the attention vector in inter-metapath aggregation, and an embedding dimension of 64.

To assess the predictive performance of graph-based node representations for task state prediction, we trained both supervised and unsupervised learning models on four datasets, using the graph output features as inputs. The query segments were divided into training, validation, and testing sets, with a split ratio of 60-20-20%, respectively.

Classification: Predicting task states

In our experiments, we employed a widely used classification model, Support Vector Machine (SVM), for classifying task states based on the generated embeddings of labeled query segment nodes. Specifically, we utilized a linear SVM classifier, where the embeddings were used as input features to classify each query segment into one of the predefined task state classes.

Clustering: Predicting task states

To explore unsupervised learning for task state prediction, we applied a clustering algorithm based on Partition Around Medoids (PAM) [133]. This technique aims to cluster data points by minimizing the average dissimilarity between each data point and its closest selected data point (medoid). To determine the optimal number of task clusters, we employed silhouette coefficients as a measure of cluster quality.

7.4 Results and Evaluation

We use Macro-F1 and Micro-F1 metrics to report our classification performance for evaluations. In addition, we use standard clustering evaluation metrics – normalized mutual information (NMI) [196] and adjusted Rand index (ARI) [308] as the metrics to evaluate the clustering performance. Finally, we report the averaged results in Table 7.3.

7.4.1 Baselines

There is currently no established large-scale state-of-the-art model for task state detection in the field of IIR. Previous works have mainly explored basic machine-learning approaches to predict search tasks. Therefore, as baselines for our experiments, we selected two previous models that utilized basic behavioral features for classification and clustering tasks. The list of baseline models is as follows:

- State prediction classification model using SVM [175].

- Unsupervised state clustering model using K-Modes [173].

Table 7.3: Task State prediction (%).

Dataset	Metrics	Unsupervised		Semi-supervised	
		K-Modes	HGN-GNN-based PAM	SVM	HGN-GNN-based SVM
TREC'14	Macro-F1	–	–	57.8	**60.7**
	Micro-F1	–	–	58.7	**60.7**
	NMI	0.59	**0.62**	–	–
	ARI	0.54	**0.68**	–	–
KDD'19	Macro-F1	–	–	57.3	**61.8**
	Micro-F1	–	–	56.2	**62.3**
	NMI	0.53	**0.67**	–	–
	ARI	0.57	**0.65**	–	–
Study 2	Macro-F1	–	–	47.3	**49.9**
	Micro-F1	–	–	50.4	**51.3**
	NMI	0.43	**0.43**	–	–
	ARI	0.42	**0.44**	–	–
Simulated	Macro-F1	–	–	62.1	**70.8**
	Micro-F1	–	–	61.3	**71.2**
	NMI	0.66	**0.72**	–	–
	ARI	0.68	**0.75**	–	–

The results presented in Table 7.3 indicate that our classification model consistently outperforms the other baselines across different training algorithms and datasets. However, the heterogeneous GNN-based embedding achieves even better performance, demonstrating that the GNN architecture, which utilizes the heterogeneous node features, significantly contributes to improved prediction performance. The performance gain obtained by our model over the baseline SVM ranges from 3-5%, indicating that graph embedding contains richer information than basic raw features.

We compared the task state clusters' quality with the annotated ground truth data to validate them. We employed two evaluation metrics, NMI (Normalized Mutual Information) and ARI (Adjusted Rand Index), to assess the clustering performance. NMI measures the mutual information between the ground truth labels and the labels predicted by the clustering method, evaluating the clustering result quality. On the other hand, ARI measures the agreement between the ground truth labels and the predicted labels, providing a similarity measure on a scale of 0-1.

The NMI and ARI values in Table 7.3 demonstrate that our clustering model consistently performs well in clustering task states. It can be observed that the traditional feature-based models do not possess significant advantages over the metapath-driven heterogeneous graph neural network embedding in clustering tasks.

7.4.2 RQ3.2 Ablation Study

To address RQ3.2 and evaluate the effectiveness of embedding features in task state detection, we conducted additional experiments using two different sets of input features. The purpose was to assess the impact of node embedding based on textual content and search behavior signals separately.

We built three classifiers and three clustering models, resulting in a total of six models for the TREC 2014 Session Track and Study 2 dataset. The models were trained using the following combinations of input features: textual content-based embedding (i.e., query and item texts), search behavioral features-based embedding, and embeddings based on all available textual content and behavioral attributes. The experiments were performed with the same settings as before, except for the abovementioned differences.

The results obtained from these model variants are presented in Tables 7.4 and 7.5. It can be observed that utilizing all features in node representations leads to significant performance improvements in both classification and clustering models, compared to the text-based and behavior-based embedding models, for both datasets. This finding highlights the importance of incorporating node-type-specific attributes and content transformations to represent nodes effectively. It also underscores the usefulness of capturing rich structural features of search sessions through metapath instances and neighborhood node information.

Furthermore, the comparable performances of the textual features-based embedding models suggest that textual features strongly influence task state detection. This observation suggests the potential benefits of applying task state detection at an early search stage, enabling applications such as state-aware task completion and proactive recommendation [175].

In addition to quantitative evaluations, we conducted a qualitative assessment by comparing the resulting clusters with the annotated state labels. The agreement between the clustering algorithm and human annotations, measured by Cohen's Kappa, was above 60%. This reasonably high level of consensus indicates that the task state typology generated by the metapath-driven heterogeneous embedding-based node clustering algorithm is reliable and can be used for further analysis.

Thus, our task state detection model, which utilizes heterogeneous node embedding, demonstrates significant performance improvements compared to equivalent models that do not utilize embedding graph features. Moreover, the model built on all textual and structural features-based embedding outperforms the models based on behavioral features-only

Table 7.4: task state prediction (%) in TREC'14 dataset.

Models	PAM		SVM	
	NMI	ARI	Macro-F1	Micro-F1
All feature-based HGE	0.62	**0.68**	**60.7**	**60.7**
Textual Content-based HGE	**0.67**	0.61	59.1	59.3
Behavior Features-based HGE	0.40	0.50	57.4	57.6

Table 7.5: task state prediction (%) in Study 2 dataset.

Models	PAM		SVM	
	NMI	ARI	Macro-F1	Micro-F1
All features-based HGE	**0.67**	0.65	61.8	**62.3**
Textual Content-based HGE	0.64	0.62	55.7	52.1
Behavior Features-based HGE	0.59	0.61	49.5	50.1

embedding and textual content-only embedding in identifying task states. These findings highlight the effectiveness of incorporating textual and structural features in the node embedding process. Furthermore, the results indicate that it is possible to predict task states by employing the textual content-based heterogeneous graph embedding alone. This finding encourages further research, exploring the potential of leveraging textual content for task state detection.

Overall, the combination of quantitative and qualitative assessments demonstrates the effectiveness and reliability of the embedding features in detecting task states, as well as the value of incorporating heterogeneous node embedding in the task state detection process.

7.5 Implications

The study focuses on the challenging problem of identifying searchers' evolving task states during the search process. Applying a heterogeneous graph embedding-based neural networks model, the study models web sessions as heterogeneous graphs of searcher-system interactions.

However, there are some limitations and constraints to the study. The lack of task state-labeled datasets prevented comparison with other existing work on task state detection. Additionally, the human-annotated task state labels may contain biases, which should be taken into consideration.

The identification of searchers' evolving task states is crucial for developing adaptive search and information retrieval systems that can support searchers throughout their tasks. The proposed heterogeneous graph-based model provides a means to understand and explain search behavior patterns in relation to tasks, search intents, and task states.

Chapter Summary

- Searchers' evolving focus of attention states can be influenced by multiple latent factors during the search process.
- The study applies a heterogeneous graph embedding-based graph neural networks model to represent web sessions and recognize task states that may shift during a session.
- The model considers semantic information of queries and latent contextual factors to interpret how searchers' task states emerge and evolve.
- Evaluation of the proposed approach demonstrates the feasibility of automatically extracting task states from search session behavioral variations.
- However, the study has limitations such as the lack of comparison with other task state detection methods due to insufficient labeled datasets and the potential bias in human-annotated task state labels.
- Identification of evolving task states is crucial for adaptive search and information retrieval systems, providing support to searchers throughout their tasks.
- The study contributes to task-based information retrieval research by generating new knowledge on the relationship between searcher's task states, search sessions, and task goals.
- Task state-based systems can lead to the development of intelligent and proactive search and recommendation systems.

- Understanding the searcher's current task state is the first step towards achieving these goals.

Chapter 8

A SYNTHETIC SEARCH SESSION GENERATOR FOR TASK-AWARE INFORMATION SEEKING AND RETRIEVAL

Prior to experimenting with our task representation model in search and recommendation models, we first tackled the challenge of dealing with small datasets for large-scale experimentations. To address this, we developed a simple workflow for synthetically generating search sessions with task state labels. This was done before addressing RQ4, and the complete study can be found in [244].

8.1 Introduction

The development and evaluation of methods for accurately extracting task states and implementing task-aware adaptive IIR systems or any IR systems on a large scale heavily rely on extensive search history logs. However, obtaining a substantial amount of real-world search-log data is challenging due to various factors such as ethical and privacy concerns. As a result, the lack of publicly available session search behavior data labeled with task state information often hinders the progress of research in exploring new directions for IIR. To overcome this issue, one approach is to artificially generate a significant volume of synthetic or simulated search session logs with task state information. In this study, we have developed a robust, scalable, and user-friendly simulation model to mimic user actions and generate a substantial amount of realistic synthetic web search sessions, specifically in a complex task scenario, along with task state information. This synthetic data can then be utilized to advance research efforts in task-based IIR systems and their evaluation.

Simulating user search behaviors poses a significant challenge in information retrieval, as evidenced by ongoing research in this field [48, 194, 193, 56, 197]. The complexity of information searching behavior, which involves various types of interactions driven by different motivations, contributes to the difficulty. Previous studies in user behavior simulation within the IR domain have examined specific aspects of the search process, such as query formulation and reformulation [12, 58], browsing and clicking on results [70, 88], as well as session identification and abandonment decisions [194]. Additionally, some simulation studies have combined subsets of these user behaviors to simulate realistic user interactions and search sessions [14, 11, 193, 310, 48].

Hence, our objective is to develop and empirically evaluate a robust and dependable model for generating large-scale synthetic sessions with task states. Our simulation model comprises multiple machine learning modules, each dedicated to modeling different aspects of user behaviors. The following research questions guide our approach:

RQ1: Can we create complete synthetic user search sessions which mimic the actual user search session behavior but require minimal human workload?

RQ2: Can users differentiate between model-generated and actual user sessions?

8.2 Synthetic Search Session Generation Process

A user initiates the search process with a task and information needed. Therefore, an ideal search session consists of a task, topic, a clear beginning point, one or multiple queries, retrieved ranked result lists, users' search interactions on result pages and viewed documents, time spent on those documents, and users' evolving task states [310]. Figure 8.1 shows a user's search sessions with various actions and decision points. Therefore, to simulate a complete search session, we need to simulate all the aspects of the session. There are a few studies that also tried to simulate search session interactions. For example, [14] simulated user sessions with various query modification strategies and created sessions with combinations of these query modification and result scanning methods. In addition, [193, 192] proposed an open-source toolkit, SimIIR, for conducting IIR experiments, which consists of several configurable components such as query generation techniques and stopping strategies. Our approach differs from existing session simulation models regarding session creation and evaluation of the created search sessions. We develop an integrated workflow of several machine-learned

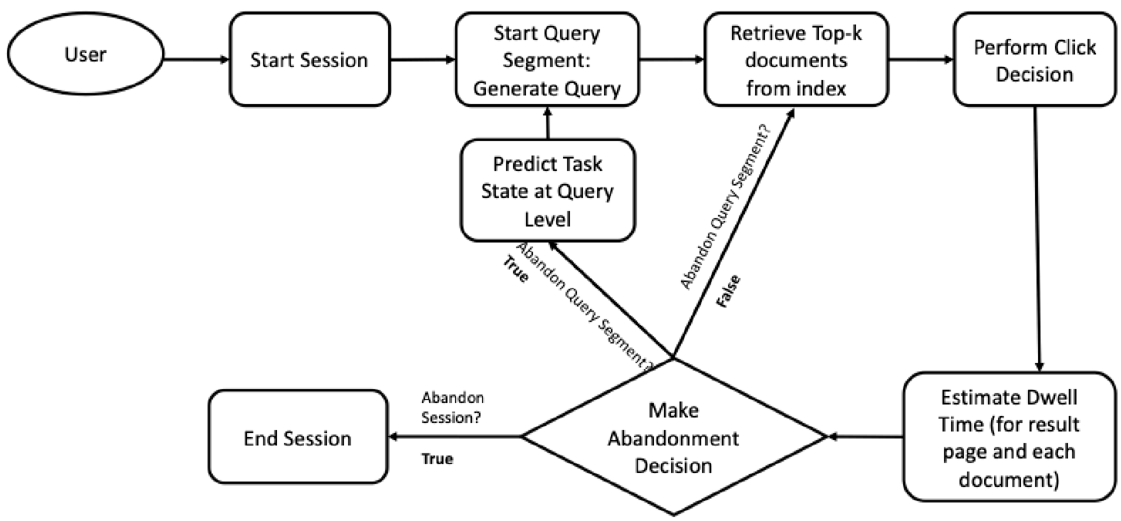

Figure 8.1: Overview of the Workflow of complete session generation process (for details, see Section 8.2).

models to generate and reformulate queries, predict user click decisions, estimate dwell time, identify task states, and make abandonment decisions. The scenario we address is inspired by typical TREC Session [57] style search sessions where search topics are given as a textual description of some information needed. Based on past studies, we define a search session as a sequence of queries with the same task topic.

8.2.1 Query Generation

Many studies aimed to generate search sessions with a sequence of queries with similar intents. Existing studies have primarily used either static rule-based approaches [15, 280] or language models based on topics, documents, and snippets [194] to generate queries. Based on the studies which found that users tend to form queries in two or three words in textual search [119], many query generation models (e.g., [193, 48]) simulate queries with three terms. We use two different approaches to formulate queries in two different scenarios. As we do not have enough search behavioral signals at the beginning of search sessions to generate the first query to start a session, we utilize a language model [98] that leverages BERT embeddings and cosine similarity to extract essential terms from available task information. Also, we experiment with three to five-term-based query lengths. After generating the first query, we use contents from users' clicked documents in previous query segments along with task descriptions to reformulate queries. We compute the maximum likelihood of terms with terms that appear in task topic descriptions, queries, and retrieve documents in previous search interactions [193].

8.2.2 SERP Scanning and Click Simulation

After retrieving documents that may be relevant to answering the information needs, we perform click predictions on the result lists. To predict whether a user clicks on a document from the retrieved document lists, some models consider past session history and all retrieved documents for the current query [69, 184, 306]. There are many well-performing click models [69]. We generate a user's click on a result snippet using methods described in [69]. The algorithm calculates the conditional probability of clicking on a document, using query-document pair similarity (especially for the first query) and aggregating users' search behaviors on the result pages for a query (e.g., clicks, dwell times). We found that a dependency model that uses features of all snippets in a results page is a better click model than one that uses only components of a target document and that using elements of the session history improves precision.

8.2.3 Dwell Time Generation

We can calculate dwell time based on users' spending time on each search action in the search process. For example, the dwell time on a retrieved document starts with a user clicking on the document listed on a retrieved result list and ends with returning to the result lists or abandoning the search. Among other models of dwell time prediction, [282] adopted a factorization model that captures the interaction between users and webpages to create a model that predicts the dwell time of a user on a certain depth of a document. To simulate dwell times at each search action, we extracted time distribution patterns from user sessions from TREC Session Search 2014 [57] and reproduced the time distribution for the generated sessions.

8.2.4 Session and Query Segment Length and Abandonment Prediction

Session abandonment has been studied extensively, such as search session satisfaction prediction and session abandonment type prediction. However, users eventually abandon a search process due to achieving their task goal(s), dissatisfaction with the results shown by the search systems, or simply for the time constraint. For session length generation, we use session, topic, and interaction features and SERPs: the title of a snippet, the text of a snippet, and the URL, and fit them into a regression model to predict whether a session or query segment was abandoned or not.

8.2.5 Task State Generation

A few models of user task states are available. [193] in their *User State Model* (USM) defines task states as "the user's cognitive state". Whereas [173] extends this idea and identifies four broadly applicable task states (described in chapter 6) in search sessions based on users' search behavioral patterns where users try to evaluate known information and their existing knowledge further. [48] defines task state and "subtopic state," which consists of a representation of the terms about the task subtopic, which then be used for query generation and click-decision. Our simulation model uses the typology defined in Chapter 6 to determine the user task state for each query segment. We conceptualize search task state as the user's focus of interest at a query segment which consists the motivation behind the query, users' perceived problem and help, and the users' search strategy (e.g., browsing and click-decisions) for that query. Therefore, the task state typology in chapter 6 would be helpful for our task-state simulation. To operationalize these states, we define each search state as a query segment with all the available contextual information associated with the segment in the search log. Therefore, query segments are clustered using K-means clustering techniques based on their content and search activity patterns.

8.3 Experimental Setup and Evaluation

As described in Figure 8.1, we generated users' search interactions to build synthetic search sessions to test our model. Our experimental setup was primarily influenced by TREC Session Search type user study setups where search tasks and topics were provided. In this experiment, we used 60 topic descriptions from TREC 2014 Session Track dataset [57]. Based on the TREC 2014 Session Track topic descriptions, we generated the search sessions by following the steps mentioned in the previous section.

To generate each component or interaction event of a search session using the individual module, we use the default configurations (base model) for each model architecture to make the process straightforward. However, our model parameters can be tuned if needed. The process starts with generating a query based on a given task description. Then top 10 documents that may answer the information needed for that query are retrieved to create a ranked list of potentially relevant results. Next, the click prediction algorithm estimated users' click actions. Dwell times on the clicked documents and result lists were also calculated. Finally, we made query and session abandonment decisions. If query segment abandonment was detected, then task state prediction was made for each query segment to understand users' latent focus for that query segment. The process is repeated until the session abandonment is decided.

8.4 Evaluation

Evaluation of the synthetic search session generation methods is the most challenging part. However, for this study, we care most about whether the sessions we generate are "useful" for evaluating task-session-based IIR systems. Therefore, to answer our RQs 1 and 2, we evaluate the quality of our model-generated synthetic sessions through a user experiment.

To measure the quality of synthetic sessions, we wanted to see whether a human searcher can distinguish synthetic from user-generated TREC sessions. In order to evaluate how users perceive synthetic sessions and task states, we conducted a small user study and compared the sessions generated to sessions from the TREC 2014 Session track. Following similar approaches from past IIR research [173, 175], we asked a set of human assessors to look at the synthetic and TREC sessions. We had to judge (i) whether a session seemed to be simulated or of human origin and (ii) task state labels. We created search log excerpts with task topics, user ids, timestamps, queries, retrieved documents, and task states by mixing 10% of both synthetic and user sessions from the TREC Session Track for randomly selected 5 task topics. We provided this search log to each human assessor and asked them to follow the sessions. The judges then reviewed the topic descriptions, examined the provided sessions, and labeled each session as human-generated (from the TREC Session) or synthetic. The result shows that, in the majority (60%) cases, assessors could not differentiate the synthetic sessions from TREC ones. However, note that a person might have guessed the session origins due to the pattern of keyword-based query generation. On the other hand, the inter-annotator agreement between our task-state predictive module and human annotators (combined both rounds) was fair (63%), which showed the usefulness of our model.

8.5 Assumptions and Limitations

Our evaluation showed that human judges could not distinguish synthetic sessions from sampled TREC sessions. Therefore, some authenticity can thus be attributed to the synthetically generated sessions. Thus we showcased the potential, but as with other attempts to generate synthetic search sessions, this work is not entirely fault-proof. At the same time, our observations during these experiments and while shaping this preliminary study showed that parameters for synthetic session generation are very dynamic, as well as task situations and topic-specific issues. Therefore, before researchers decide to use the workflow in their research, it is worthwhile to rethink the assumptions behind this work that might have influenced the data generation process and consider some of the limitations of this work. For example, this study did not consider navigational, known-item, or fact-finding tasks that often generate short sessions with limited queries. Furthermore, the working definition of the session used in this work simplifies the modeling of search behavior, as it is agnostic of temporal criteria regarding how long a user has been idle between sessions. Also, some of the parameters used in the model, such as dwell times and session lengths, are based on set values found in TREC Session Search scenarios which may not be appropriate in other complex task scenarios. Addressing the outlined shortcomings is an interesting direction for future work, which we plan to continue pursuing and updating our model.

Chapter Summary

- Authentic search sessions with task states can be simulated using a step-by-step mixed model workflow by employing different strategies of simulating different aspects of user web search sessions.

Chapter 9

RESEARCH QUESTION 4: APPLY CONTEXT-INDEPENDENT TASK REPRESENTATIONS INTO SEARCH PROBLEM

The first experiment focuses on showcasing the utility of the extracted task information (Chapter 7) in diverse search and recommender applications across different domains. Specifically, we investigate the potential of leveraging task information in re-ranking models and search personalization techniques. Our objective is to demonstrate that incorporating task information proves more beneficial compared to conventional features typically employed in retrieving and ranking information objects that align with users' requirements.

9.1 Search Ranking and Re-ranking

The ranking is crucial in information retrieval and is fundamental to various IR-related applications, including search result ranking, query auto-completion, and video recommendation. The primary objective of ranking models or algorithms is to arrange or sort a collection of documents, denoted as d that match a given query denoted as q, based on a predefined criterion. The aim is to ensure that the most relevant results appear prominently in the list of results presented to the user. Learning to rank (LTR) is an approach that utilizes supervised learning techniques to construct re-ranking models for IR systems. This approach leverages a training dataset comprising queries, documents that match these queries, and corresponding relevance scores. By employing this training dataset, a learning-to-rank algorithm develops a re-ranking model capable of predicting the relevance of documents in the collection d for a given query q.

9.2 Experimental Setup and Evaluation

To tackle the challenge of utilizing our context-agnostic task representation to generate a ranked order of documents, we employ task embeddings derived from our model in a sequence-wise learning-to-rank (LETOR) model. This LETOR model can score and sort a sequence of ranking candidates simultaneously [284]. LETOR represents a specific ranking approach that employs supervised learning techniques to construct ranking models and establish an order for a list of ranking candidates. In the usual scenario, a LETOR algorithm learns a scoring function using a training set comprising queries, a set of documents that correspond to these queries, and their respective relevance scores. This scoring function assigns a score to each document, enabling further prediction of document relevance in a document collection for a given query. The documents are then ordered in descending order based on these scores [315].

Based on the listwise learning to ranking model ListNet by [50] and SERank [284], the formal problem statement of our approach is as follows: (In following descriptions, we use a subscript to indicate both the index of queries and the index of documents for a specific separated by a ”,”.)
Suppose $Q = \{q_1, q_2, q_3, ..., q_m\}$ is the set of all queries and q_i is the i^{th} query within the set, $T = \{t_{i,1}, t_{i,2}, t_{i,3}, ..., t_{i,l}\}$ is the set of task contexts for each q_i, $D_i = \{d_{i,1}, d_{i,2}, d_{i,3}, ..., d_{i,n_i}\}$ is the set of candidate documents or items associated with the q_i where n_i denotes the size of D_i, and $d_{i,j}$ denotes the j-th document in D_i. Individual list of documents D_i are associated with a list of labels or judgment scores $y_i = \{y_{i,1}, y_{i,2}, y_{i,3}, ..., y_{i,n_i}\}$ where $y_{i,j}$ denotes the relevance or judgment of the document $d_{i,j}$ with respect to q_i. $y_{i,j}$ is the relevance judgment score for j-th document in d_i during a given q_i. The output of the model would be list of ranked relevance score of a given q_i and a document d_i in the list.

A feature vector containing features and relevance labels $y_{i,j}$ for each document $x_{i,j} = \varphi(q_i, d_{i,j})$ is created from each document pair $(q_i, d_{i,j}), i = 1, 2, ..., m; j = 1, 2, ..., n_i$, where φ

denotes the feature functions of a document pair. A list of features $x_i = (x_{i,1}, x_{i,2}, ..., x_{i,n_i})$, and a list of corresponding scores form a training set $y_i = (y_{i,1}, y_{i,2}, y_{i,3}, ..., y_{i,n_i})$ form an instance in the training dataset $S = (x_i, y_i) \lim_{i=1}^{m}$. For the individual feature vector $x_{i,j}$ corresponding to document $d_{i,j}$, a ranking function f is generated which outputs a score $f(x_{i,j})$. For a list of feature vectors x_i a list of scores $z_i = (f(x_{i,1}), ..., f(x_{i,n_i}))$ can be obtained. The objective of the training ranking model is to learn a scoring function, which maps the input feature vector of x_i to an output relevance value R_i, and define a score function that minimizes the empirical loss over the training set. The total losses with respect to the training data obtains a listwise ranking or loss function $\mathcal{L}$ to minimize the loss: $\sum_{i=1}^{m} L(y_i, z_i)$. The documents $d_{i'}$ may be ranked in descending order of the scores in a listwise manner.

Several probability models can be employed to calculate the listwise loss function. In our approach, we utilize a learning method that optimizes the listwise loss function based on the probabilities of the top-k or top-one documents. This learning method involves neural network modeling and utilizes Gradient Descent as the optimization algorithm, inspired by the ListNet approach [50].

For the neural network ranking function, denoted as ω, we represent it as f_ω. This ranking function takes a feature vector $x_{i,j}$ as input and produces a score, denoted as $f_\omega(x_{i,j})$, for each document. Thus, given a query q_i, the ranking function f_ω generates a score list $z_i(f_\omega) = (f_\omega(x_{i,1}), f_\omega(x_{i,2}), ..., f_\omega(x_{n_{i,j}}))$, where $n_{i,j}$ represents the total number of documents for query q_i.

To calculate the top-k probability for documents $(d_{i,j_1}, d_{i,j_2}, ..., d_{i,j_t})$, we consider the probability that each document appears among the top k documents. The specific calculation of this probability depends on the chosen probability model and optimization method used in the learning process:

$$P_{z_i(f_\omega)}(x_{i,j}) = \frac{exp(f_\omega(x_{i,j}))}{\sum_{k=1}^{n_i} exp(f_\omega(x_{i,k}))} \tag{9.1}$$

With Cross Entropy as metric, the loss for q_i can be measured as follows:

$$\mathcal{L}(y_i, z_i(f_\omega)) = -\sum_{j=1}^{n_i} P_{y_i}(x_{i,j}) log(P_{z_i(f_\omega)}(x_{i,j})) \tag{9.2}$$

where y_i is a judgement value for an item with respect to a feature ; P_{y_i} is a probability of y_i; z_i is a list of scores for y_i with a neural network model function f_ω applied to individual y_i according to an equation $z_i = (f(x_{i,1}), ..., f(x_{i,n_i}))$. $P_{z_i(f_\omega)}$ is a probability of z_i; and $-\sum_{j=1}^{n_i} P_{y_i}(x_{i,j}) log(P_{z_i(f_\omega)}(x_{i,j}))$ is the cross entropy for the q_i. $j = 1$ indicates that each element in $x_{i,j}$ is an element of collection q_i – the top k subgroup of the retrieved list of documents.

9.2.1 Dataset

In our experiment, we aggregate all the session logs utilized in Chapter 7 to obtain a substantial number of search iterations involving query reformulations, thereby simulating the completion of a complex search task. The combined session logs encompass over 3000 search sessions, covering diverse topics. Each session comprises a sequence of queries and reformulations. The log data includes information such as the number of clicked documents, snippets copied by the user, and the user's dwell time on each document.

For evaluation purposes, we create a test set consisting of 10 relevant documents for each query, along with their corresponding relevance scores. This test set serves as a benchmark for assessing the performance of our models and algorithms.

Metrics

We use standard ranking evaluation metrics from the literature to evaluate our models. We consider all the documents corresponding to a query q as the candidate document set for q and the relevance scores and item ranks of the document as the ground truth of recommendation. The metrics we used include the Mean Average Precision (MAP), the

Table 9.1: Performance on the Combined Dataset

Model	nDCG10	MAP10
Baseline with LETOR	0.46	0.63
Our model generated task vectors	0.48	0.68
All featues	**0.51**	**0.72**

average multiplicative inverse of the correct document's rank, and Normalized Discounted Cumulative Gain (nDCG).

Competing Models and Ablation Step

The metrics evaluate ranking models for the first top-ranked results, for example, the proportion of the top 10 results relevant on average over many queries. We generate three ranking models based on three types of feature sets. Our baseline model used the standard LETOR features for learning to rank techniques [219], which do not consider task-related features. The second model uses the task embeddings derived from our model, and the third model contains all features, including the standard LETOR features and the task embeddings from the model we built.

9.3 Results and Discussion

To guide our analysis of the results, we formulate the following research question: **RQ:** *Can the combination of task-related feature vectors effectively enhance search performance?* The experimental results demonstrate that our task-aware LETOR model outperformed the baseline model. Specifically, based on the nDCG scores for the top 10 document rankings, our model outperformed ListNet with standard LETOR features. As shown in Table 10.2, utilizing task embeddings improved ranking performance effectively. This outcome indicates that incorporating task context features can enhance support for searchers dealing with complex search tasks.

The results above indicate the successful performance of task context representation. This outcome can be attributed to the fact that multiple features have the ability to accurately represent the context of a search task better than a single feature alone. However, it is essential to note that each feature typically contains noise and uncertainty. Considering these sources of uncertainty, common in real-world applications, we observed that individual contextual features alone could not sufficiently improve ranking. Nonetheless, combining multiple task aspects makes it possible to create a precise representation of task context for the search process, thereby providing more effective contextual features for ranking models.

9.4 Implications

This study aimed to integrate our task state embedding into learning-to-rank algorithms to optimize document ranking for a given query, thereby providing reactive, proactive, and dynamic support. Our approach has yielded insightful findings. The ranking results demonstrate a strong correlation between searchers' evolving and overall task contexts and their search behaviors, selection of search strategies, and relevance judgments during different stages of a search session.

Chapter Summary

- The study extracts dynamic task information as HGNN embeddings and incorporate them into learning-to-rank algorithms.
- The study reveals that searchers' evolving and overall task contexts are closely linked to their search behaviors, selection of search strategies, and relevance judgments within various sub-steps of a search session.
- Although the results show moderate improvements, this study serves as a prelimi-

nary investigation into the effectiveness of incorporating task contexts to enhance search systems' capability in supporting complex search tasks.

Chapter 10

RESEARCH QUESTION 4: APPLY CONTEXT-INDEPENDENT TASK REPRESENTATIONS TO GENERATE RECOMMENDATIONS

In the second experiment, we implemented our task model into a recommendation problem.

10.1 Introduction

People use search and recommendation systems not only for easy access to information but also for accomplishing complex everyday tasks that may require multiple rounds of searching and browsing information of various kinds. In order to help users in fulfilling their overall task at hand rather than supporting only one task at a time, search and recommendation systems must address users' shifting task goals [260, 269].

10.2 Recommendation Model using Task State Model

We build a task-based recommender model with our HGNN-based task embedding model on top of a collaborative filtering approach in order to return top-N recommendations for users based on their tasks. This model contains two layers: Layer 1 is the heterogeneous GNN layer that models the relationships between various information items and users and represents users' task states. This GNN layer uses a graph structure to represent tasks based on the relationship among graph nodes (e.g., sessions, users, and items). Finally, layer 2, a recommendation layer, uses a collaborative filtering algorithm to return top-N recommendations for each user based on their behavior and tasks.

10.2.1 Task Embedding Layer

We explain our task representation layer in Chapter 7.

10.2.2 Recommendation Layer

In step 2 of our approach, we utilize the outputs generated by the task embedding layer as inputs to a Collaborative Filtering (CF) model in order to obtain a list of the top 10 items that are potentially relevant to a user's task. Collaborative filtering operates on the principle of identifying similar users based on their historical interactions with items and recommending items that have been favored by those similar users. Essentially, collaborative filtering assumes that users who have exhibited similar preferences in the past are likely to continue having similar preferences in the future. Hence, the system assesses the similarities between users or items, arranges the recommendations accordingly, and returns the top-N items to the user.

To summarize, we use the HGNN-based task representation module to extract sequential user-item interactions for sessions by metapaths to develop task representation. In the next step, we use a CF-based recommendation module in order to extract top-N recommendations based on those task embeddings.

10.3 Experiments

This section outlines the specifics of our experimental setup. To implement the two layers mentioned above, we established a learning rate of 0.05, a dropout rate of 0.01, and utilized 2 layers over 100 epochs.

Table 10.1: Statistics of datasets

Statistics	UB	Tmall
# of Users	987,994	17,209
# of Items	4,162,024	16,176
# of Interactions	100,150,807	831,117

10.3.1 Datasets

We conduct experiments on two datasets that are commonly used in recommendation studies. These datasets are (1) **User Behavior (UB)** [2]: this dataset includes the users' purchasing behavior of Taobao, which Alibaba offers. These behaviors are click, purchase, add to favorite, and add to shopping cart from November 25 to December 03, 2017, and (2) **Tmall** [3]: this dataset contains anonymized users' shopping logs of Tmall.com in the past six months before and on the "Double 11" day. This dataset includes different action types like click, purchase, add to favorite, and add to cart. We allocate 30% of these datasets for the train set, and rest for validation, and test sets. The statistics of these datasets are shown in Table 10.1. Since the datasets do not contain any task state labels, we use the unsupervised clustering techniques to cluster the task states.

10.3.2 Evaluation

To assess the performance of our model, we employed two widely used ranking performance metrics: Normalized Discounted Cumulative Gain (nDCG@K) and Mean Reciprocal Rank (MRR@K).

nDCG@K measures the effectiveness of a recommendation algorithm by comparing the

[2] https://tianchi.aliyun.com/dataset/649

[3] https://tianchi.aliyun.com/dataset/42

ranked list of recommended items to a list of ground truth items that the user is known to be interested in. It takes into account the position of each relevant item in the ranked list and assigns higher scores to items that are ranked higher. In our evaluation, we considered $K = 10$, which means we evaluated the performance of the model based on the top 10 recommended items.

MRR@K, on the other hand, calculates the average reciprocal rank of the relevant items in the ranked list of recommended items. It provides a measure of how quickly a relevant item is found in the list. For our evaluation, we set $K = 5$, which means we focused on the performance of the model in terms of the top 5 recommended items.

By utilizing these metrics, we were able to quantitatively assess the effectiveness of our model in recommending relevant items to users based on their tasks.

10.3.3 Baselines

In this study, we have chosen the following baseline models for task-based top-N recommendation in sessions:

- NMF (Non-Negative Matrix Factorization) [151]: This algorithm recommends a set of N items to a user based on their historical preferences. It achieves this by factorizing the user-item interaction matrix into two non-negative matrices, representing user and item factors. NMF leverages these factors to make personalized recommendations.

- GRU4REC [109]: This model takes a sequence of items as input, representing the items interacted with by the user in each session. GRU4REC utilizes a Gated Recurrent Unit (GRU) architecture to predict the probability of the next item in the session. Recommendations are then made by selecting the items with the highest predicted probabilities.

Table 10.2: Performance of our task-based model along with the t-scores for each model. Significant values at $p < 0.05$ are marked with '*'.

	UB			Tmall		
Model	nDCG10	MRR5	t	nDCG10	MRR5	t
MF	0.45	0.48	-	0.47	0.45	
GRU4REC	0.27	0.25	**1.42***	0.25	0.23	**1.07***
SR-GNN	0.25	0.13	-1.45	0.20	0.22	0.91
Our model	0.49	0.50	**-2.80***	0.51	0.48	**4.51***

- SR-GNN (Session-based Recommendation with Graph Neural Networks) [300]: This algorithm learns session representations using a graph neural network. It predicts the probability of each item being the next item in the session. Items are ranked based on their probability scores, and the top-N items are recommended to the user. For training the SR-GNN model with our data, we utilized the code repository provided by the authors of the original paper.

  To train SR-GNN model with our data, we used the code repository provided by the authors of the above-mentioned paper[4].

These baseline models were selected to compare and evaluate the performance of our proposed approach for task-based top-N recommendation in sessions.

10.4 Results and Discussion

To organize the outcome of this study, we answer the following research question based on the analysis and results:

[4]https://github.com/CRIPAC-DIG/SR-GNN

RQ: *How can a GNN-based task state model effectively personalize task-based recommendations in each session?*

We conducted multiple runs (10 times) of our proposed model and the baseline models in our experiments to obtain reliable results. The average outcomes were then calculated and compared. As shown in Table 10.2, the experimental results indicate that our task-embedding-based recommender model outperformed the three baseline models. Specifically, based on the nDCG and MRR scores for the top-N recommendations, our model demonstrated superior performance compared to SRGNN and GRU4REC. These findings highlight the effectiveness of incorporating task context features in supporting users with complex tasks.

Additionally, we conducted a two-sample t-test to assess the performance of each model compared to our fundamental baseline model, NMF. The results showed non-significant improvements for the SR-GNN model compared to the first baseline model. It is noteworthy that SR-GNN did not perform poorly in these experiments. One potential reason behind this observation could be the relatively smaller size of our datasets, as dataset size can influence performance. Nonetheless, the results demonstrate the benefits of leveraging task context to personalize recommendations based on users' tasks.

Our findings support the notion that task context contributes to improved performance. This is likely because multiple features can more accurately represent the search task context than individual features, which are often noisy and uncertain. Given the common occurrence of uncertainty in real-world applications, relying solely on individual contextual features to enhance recommendations is challenging. However, combining multiple features can create an accurate representation of the task context, providing more effective contextual features for recommendation models.

Chapter Summary

- Incorporating dynamic task features into recommendation algorithms to optimize top-n recommendations and assist users in completing tasks.
- Visualization of the user-system interaction process as an information graph, with users and items represented as vertices.
- Utilization of heterogeneous graph neural network embedding-based task modeling, incorporating contextual information from user sessions.
- Training of a task-aware collaborative filtering model based on inferred tasks from user-system interactions.
- Evaluation of the methods using publicly available log data, demonstrating superior performance compared to baseline recommendation models.
- Close association observed between users' evolving task stages, overall task contexts, browsing behaviors, and relevance judgments in sub-steps of a session.

Part III

CONCLUSIONS

This final part of the thesis summarizes the findings from this research with assumptions and limitations and explores several potential areas for future work.

Chapter 11
CONCLUSION

Our dissertation focuses on analyzing users' search behavior, identifying their common challenges, and determining their preferred solutions. We developed techniques to extract and utilize search task information to better understand users' needs. By studying users' information-seeking behavior and characterizing the task state framework, we were able to extract tasks as a sequence of task states. We also used task information to create user models and improve document ranking and collaborative filtering for recommendations.

In this chapter, we discuss some of the implications of the research presented in the different chapters of this thesis.

11.1 Implications of the Results for Task-Aware Adaptive Support for Complex Task Doers

At the beginning of this thesis (Chapter 1) we discuss possible ways for systems to support users to complete complex search tasks: that is, the systems need the ability to capture, characterize, and execute that underlying tasks at various levels, a way to incorporate such a representation into search, recommendation applications and provide appropriate intervention support for struggling users at their time of need. The four broad research questions try to find ways to address the three objectives that we set at the beginning of the thesis. Here, we discuss how well we address the objectives stated at the beginning and how they have implications for broader task-aware adaptive supports to users.

For better understanding, we mention the objectives and research questions from Chapter 1.

Objective 1: Developing a Conceptual Model - The first objective is to develop a conceptual model that comprehends how different tasks elicit specific information needs, leading to diverse methods, strategies, and information sources for seeking information. The aim is to understand the users' tasks, the motivations behind task needs, their search approaches, and the challenges they encounter during each step of their search sessions [259, 243]. In other words, we aim to construct a task-information need-strategy-problem map, which can be utilized to offer task-based support in various information formats (e.g., suggesting queries, documents, or relevant individuals) to address the encountered problems and guide users towards successful task completion.

Objective 2: Understanding Evolving Task States - The second objective is to explore how the knowledge of evolving task states, underlying search needs, and the overall goal can enhance a system's offerings to its users. Addressing the aforementioned challenges necessitates automated methods for the sense-making of tasks and the development of task-based information retrieval systems. These systems should excel in modeling, identifying, and extracting tasks while also supporting users' complex search tasks across various search situations, modes, and modalities. Building upon the insights gained from the first objective, this dissertation aims to develop computational models that can extract the nature of users' tasks from their search behaviors.

Objective 3: Leveraging Task Knowledge in Search Applications - The third objective focuses on utilizing the acquired task knowledge in different search applications. Specifically, the emphasis is on learning task-based responsible design for scalable and efficient information systems. These systems aim to align with users' task goals and deliver relevant information in various formats, such as queries, documents, and people. By understanding and leveraging task information, this research aims to address the challenges users face during the search process and provide effective support to overcome them.

High-level Research Questions

RQ1. To what extent do various task types trigger distinct information needs, prompting diverse approaches and strategies in the search for different types of information and sources?

RQ2. What problem(s) do searchers encounter while performing a specific task, and the type(s) of help do they prefer to get from the system?

RQ3. What approaches can be employed to effectively represent and express tasks in a manner that is agnostic to the search context?

RQ4. In what ways can we integrate context-agnostic task representations into current search and retrieval approaches to enhance their efficacy?

11.1.1 Conceptualizing an Holistic Picture of Task, Applications of Our Findings and Potential Future Studies

Understanding Implicit and Explicit Search Behaviors

Study 1 (Chapter 4) **Study 2** (Chapter 5) demonstrate the importance of adopting a holistic perspective on the information search process to gain a comprehensive understanding of users' task context and motivations through their implicit, explicit search behaviors and provided feedbacks. The findings reveal that users' perception of task difficulty, task descriptions, and familiarity with similar tasks play a crucial role in determining their task state, preferred information sources, search strategies, and preferred help.

The studies show that understanding the interplay between users' implicit and explicit search behaviors provides valuable insights into their underlying information needs and search strategies. This understanding can be instrumental in designing interventions that effectively support users in addressing their specific problems.

For instance, the study participants reported a tendency to utilize multiple information sources, particularly when facing cognitive and informational needs. This highlights the importance of search and information retrieval systems offering support for accessing and integrating information from various sources and channels, including both impersonal sources (such as web pages and databases) and interpersonal sources (like social networks and forums). The system should enable seamless transitions between these sources and tailor recommendations based on the user's contextual needs and preferences for personal or cognitive information requirements.

Additionally, the studies reveal a preference for collective and collaborative information seeking on social networking sites. To support such tasks, it is crucial to incorporate features that encourage and facilitate collective information seeking. This can include providing recommendations based on popular sources, enabling sharing and collaboration among users, and leveraging collective knowledge to enhance search results. Considering users' motivations and expectations, personalized recommendations and interventions should be provided, taking into account their goals, preferences, and expectations. Doing so can enhance the user experience, leading to a higher likelihood of finding relevant and satisfactory information.

The design of the system should facilitate access to relevant sources and channels based on the specific information being sought. The presented mapping of tasks, information needs, search strategies, and problem-help relationships can be valuable in achieving this goal (Chapter '3). As users often struggle to articulate their needs, the system should assist them in identifying the most suitable sources based on the type of information they seek and their specific requirements. Providing guidance on source selection and offering options that effectively fulfill the user's needs can significantly enhance the search experience.

Future Research Directions

To gain a deeper understanding of user behavior, it is imperative to conduct extensive research that takes into consideration various contextual factors and the emergence of new technologies. It is also essential to give due consideration to user feedback as well as cultural differences. In the previous chapters, namely 4 and 5, we have collected both implicit and explicit feedback from users to obtain valuable insights into their task context. Nonetheless, further investigation is required to determine the trade-offs between analyzing implicit and explicit feedback in different task situations. This knowledge can be extremely beneficial in allocating resources for collecting feedback, which can ultimately provide better support to users during their search processes.

To gain a comprehensive understanding of tasks at different levels, it is crucial to delve into a variety of attribute sets and then combine them. This approach allows for a more nuanced analysis and provides a more complete picture of the task at hand. By exploring different sets of attributes, one can identify patterns, relationships, and dependencies that may not be apparent when examining individual attributes in isolation. Therefore, it is important to adopt a holistic approach when analyzing tasks and consider multiple attribute sets to arrive at more accurate and useful insights.

Furthermore, longitudinal studies that track the evolution of implicit and explicit search behaviors are necessary. Search behaviors are not static and can change and develop over time. By conducting longitudinal studies that observe users' search behaviors over extended periods, valuable insights can be gained regarding the transformation, adaptation, and development of implicit and explicit search behaviors and how these behaviors impact task performance. Such studies can effectively identify patterns, trends, and potential interventions to support users throughout their information-seeking journeys.

Task and User Characteristics in Modeling Users

The influence of task complexity and user characteristics on search behaviors, perceived problems, and preferred help highlights their significance in modeling users. Considering these factors when providing interventions is crucial as it ensures they are relevant and effective.

The findings from Study 1 (Chapter 4) and Study 2 (Chapter 5) hold practical value for search systems as they offer opportunities to mitigate potential issues and improve design to alleviate cognitive loads in complex task sessions. By gaining insights into the preferences and struggles of different user groups based on their inherent problem-help preferences (for example, Study 2 shows that participants who spend more time on web pages often struggle to articulate their information needs and thus prefer help in query and document suggestions), personalized models can be developed to cater to the unique needs of these groups. Furthermore, the insights regarding the heterogeneity observed at the user, interest, problem, help, and task state levels in search behavior indicate the importance of shifting the focus from entire search sessions to dynamic search task states as the primary unit of consideration and analysis. Study 2 demonstrates that users perceived problems and preferred help change as their task state evolve during the search process. Therefore this shift enables a more nuanced understanding of users' information-seeking processes and facilitates the development of targeted interventions tailored to specific task states.

Future Research Directions

Future research can focus on developing context-aware interventions based on the understanding of task and user characteristics and our unified task state representations. These interventions would adapt dynamically based on the current task context, user expertise, and familiarity with similar tasks. By leveraging real-time user interactions and contextual cues, the interventions can provide timely and personalized support to address users' specific challenges and improve their search experience.

Mapping User Behaviors to Predict Problems

We gain valuable insights by categorizing the constraints and barriers perceived by users during online information search based on their implicit search behavioral signals and explicit feedback in Study 2. Mapping users' implicit and explicit search behaviors to their perceived problems enable us to predict and anticipate the challenges they may encounter at different stages of the search process. This predictive capability holds significant importance as it allows for timely interventions that can provide relevant support precisely when users are likely to face difficulties. Such interventions can be customized to address specific pain points, resulting in a proactive approach to identifying user challenges and anticipating potential obstacles.

For instance, we can consider a scenario where a user spends an extended amount of time on search result pages without clicking on any results - a clear indication that they struggle to trust the information presented on the SERP. Leveraging our problem-help mapping, we can infer this issue and devise an appropriate intervention. The system can suggest query enhancements or refinements by displaying a notification with alternative keywords or related terms, aiding the user in better expressing their information needs. Additionally, trustworthiness indicators can be introduced, such as credit scores or ratings for web pages or sources, based on established criteria like domain authority, author expertise, or user reviews. These indicators address the user's lack of trust in retrieved information. Furthermore, considering the user's expressed concerns, interventions can provide educational resources and tips on evaluating the credibility of online information, including explanations of fact-checking techniques, source verification methods, or guidelines for identifying reliable sources.

Moreover, an ongoing feedback loop becomes essential to improve the effectiveness of the intervention strategies continuously. The system can actively solicit user feedback to gauge the provided interventions' usefulness and impact. This feedback mechanism ensures the system learns from user responses, allowing for iterative enhancements and refinement of interventions over time. By prioritizing user feedback, we can optimize the support and enhance the overall search experience.

Future Research Directions

Conducting additional user studies is imperative to establish the effectiveness and user acceptance of the abovementioned interventions. By collecting feedback from the users, we can gain insights into the relevance, usefulness, and impact of the interventions. This feedback will be invaluable in refining and improving the interventions to better meet user needs.

Furthermore, conducting comparative evaluations involving different intervention strategies is crucial. By comparing various approaches, we can assess their impacts on user satisfaction, task completion rates, and information quality. This comparative analysis will enable us to identify the most effective intervention strategies and determine which ones yield the best outcomes.

Through these user studies and evaluations, we can gather quantitative and qualitative data that will provide a comprehensive understanding of the interventions' effectiveness. This information will guide us in making informed decisions about refining the interventions and implementing improvements to enhance user satisfaction and overall search performance.

Momentary Cognitive Focus of Attention: Task Representation Model

Our research findings highlight the need to improve the support provided by existing search systems for complex and exploratory search tasks, which often require multiple rounds of interactions to complete. To address this, we propose the concept of evolving user cognitive focus or task state (Chapter 6), which involves understanding and addressing users' immediate needs and aligning with their current problematic state during the search process. By combining perceived problems, help, actions, and users' motivation, we can conceptualize tasks at a granular level as this cognitive focus state, enabling the design of interventions that cater to users' specific needs at a given time.

Search systems can leverage the insights from recognizing task states to provide adaptive support. As the task state changes, the system can dynamically adjust its behavior and suggestions to meet the evolving needs of the searcher. For example, if the system detects that the searcher has transitioned from exploratory research to focused information gathering, it can prioritize presenting more specific and relevant search results. For example, suppose a user engages with a search chatbot to explore a broad topic. In that case, the system can recognize this as an exploratory research task and provide a list of related keywords and suggestions for refining the search based on popular subtopics. The system can dynamically adjust its results and support accordingly by identifying shifts in the task state.

Another important outcome of our research is extracting and understanding user tasks (Chapter 7), which can be applied to conversational contexts. An intelligent task-aware conversational module can be developed by enriching the current conversational context with domain-specific task information. This opens up possibilities for constructing intelligent verbal responses by incorporating user context, task domain knowledge, and the user's current information needs. Task awareness in chatbots is a promising area for both research and practical applications, especially considering recent developments in generative AI.

In Chapters 6 and 7, we present a conceptual model of task states at the most granular level, which includes users' cognitive intent, problematic state, and strategies. To capture these task states, we employ a GNN-based representation model that clusters both observed and unobserved behavioral signals from users. The heterogenous GNN can capture the structural features of interactive search sessions and represent tasks based on users' behavioral signals. If we consider tasks as a hierarchical construct (as proposed in [258]) that combines macro and micro tasks, our task model can be applied to represent tasks at different levels. This is achieved by aggregating the relationships among different attributes present in the search sessions and assigning different weights to each attribute. By doing so, the model can decompose exploratory and complex tasks into smaller goals, thereby reducing cognitive load and focusing on momentary cognitive attention. The application of our task model also enables the narrowing of assistance provided to the specific state of the task at hand. This can be represented in a semantic space or an abstract space, facilitating a better identification of the task and a more comprehensive understanding of the user's underlying goals and intentions. Overall, our task model provides a flexible framework for representing tasks at various levels and capturing the dynamic nature of user tasks in the search process.

The implications of our unified and context-agnostic task representation extend beyond web search to various digital user-facing services. Understanding the task a user is trying to accomplish can improve task-aware query suggestions, document retrieval, strategy suggestions, personalization models, and recommendations. Proactive task-aware recommendations have the potential to shift web searches from reactive interactions to proactive interactions.

Future Research Directions

In future work, we plan to conduct multiple experiments to investigate the abovementioned applications. We also intend to expand the set of attributes and signals incorporated into the task model to enhance task state identification. Advanced language models such as ChatGPT can be leveraged to analyze users' textual inputs and search behavioral signals, enabling better identification of user intent, motivation, and problems. Additionally, exploring cross-device task support is an avenue worth exploring.

Advancements in large language models provide opportunities to enhance task state representations through improved textual content analysis, sentiment analysis, and entity detection within natural language processing. By incorporating insights from advanced language models like ChatGPT, our task model can better identify user intent, motivation, and problems, leading to more effective support for search task completion. It can support all search strategies identified by Belkin et al. [26, 253]. This opens up new possibilities and paves the way for advancements in the field.

Personalized help and intervention

The findings of this dissertation have the potential to significantly enhance the search experience and provide personalized support for individual users or groups of users with similar needs and challenges. By gaining insights into patterns of search behaviors in relation to task states and search intents, search systems can optimize the overall search process. They can offer customized recommendations, suggestions, and guidance that align with the current task state, thereby improving the efficiency and effectiveness of users' search activities.

One key finding is that users' desired assistance is not solely determined by their immediate perceived problems. Instead, it is influenced by the strategies they employ throughout their entire search session. This highlights the importance of personalized and adaptive interventions that take into account users' historical search behaviors. By considering each user's unique patterns, preferences, and previous interactions, search systems can dynamically adapt their support to deliver a more personalized and targeted search experience.

By mapping task-need-problem-help-strategy relationships, it becomes possible to categorize users into different types based on their shared problems, preferred types of assistance, or strategic approaches derived from their historical behavioral data. This enables the tailoring of similar support to users with similar profiles, leveraging the insights gained from their collective experiences.

Future Research Directions

In this thesis, our approach primarily relied on behavioral data obtained from medium-scale user studies, search logs, and simulated data to develop and implement task models. However, for future research, it is crucial to expand our investigations by collecting and analyzing real-world task data in naturalistic settings.

Collecting and analyzing real-world task data will allow us to validate and refine our existing task models. It will provide insights into the nuances and intricacies of users' task-related behaviors, enabling us to further improve the accuracy and effectiveness of our models. Additionally, studying real-world tasks can uncover novel patterns and dynamics that may not be fully captured in controlled experimental or simulated environments.

Implementing Task Models into Search and Recommendation Algorithms

Chapter 9 and Chapter 10 present an implementation approach for incorporating task models into search and recommendation algorithms, demonstrating the significance of integrating task contexts to enhance the capabilities of search systems in supporting complex search tasks. These findings underscore the importance of further research and refinement in order to achieve even better performance and more effective interventions. Integrating task models into search systems makes it possible to provide reactive, proactive, and dynamic support throughout the search process. Reactive support involves addressing immediate user problems or difficulties based on their current task context, allowing for timely interventions tailored to their specific needs. Proactive support goes a step further by anticipating potential challenges that users may encounter, enabling preemptive interventions to mitigate these challenges. Dynamic support ensures that the system adapts and evolves alongside users' evolving task contexts, providing continuous assistance throughout their search sessions.

However, the experiments also reveal the need to consider additional factors such as users' cognitive, physiological, and affective traits (e.g., knowledge base, search skill, emotional states) as they can influence search behaviors and task formulations. Integrating these factors into the system makes it possible to provide more accurate and personalized interventions that align with users' unique characteristics and requirements. This highlights the importance of considering a holistic understanding of users when designing and implementing task-focused search systems.

Future Research Directions

In the future, our research aims to expand the implementation of task models to various other search and recommendation problems. These include tasks such as document ranking, query suggestions, addressing cold-start problems in recommendations, and more. By incorporating task models into these areas, we can further enhance the performance and effectiveness of search and recommendation systems.

Overall, these implications emphasize the importance of leveraging users' implicit and explicit search behaviors, understanding their perceived problems, and providing personalized interventions throughout the search process. By doing so, interventions can effectively address user challenges, improve search outcomes, and enhance the overall user experience. By focusing on the integration of task models and personalized interventions, we aim to create search and recommendation experiences tailored to individual users' unique needs and preferences. This research direction holds significant potential for advancing the field and contributing to the development of more intelligent and user-centric search systems in the future.

11.2 Limitations and Contributions

While the methods and results presented in the dissertation contribute to our understanding of users' tasks, there are limitations that warrant further investigation in future research.

Regarding Study 1, it is important to note that the findings are based on an online survey conducted on a crowdsourcing platform, which restricts direct communication between researchers and respondents. The scenarios presented were simulated, and participants had to hypothesize the scenarios instead of actually performing the searches. Therefore, the responses may not reflect self-motivated behavior, and only selection choices were collected without capturing explanations behind those choices.

Additionally, self-reported data, including users' surveys on information seeking, can be subject to biases. Participants' stated actions in a scenario may not align with their actual behavior during the information search. While the findings indicate that people prefer using different information sources based on their needs, the interpretation of these needs can vary based on individuals' situational contexts.

Despite these limitations, the study reveals meaningful patterns regarding individuals' information needs, sources, and methods. Notably, people tend to utilize multiple sources to seek affirmation or alternative opinions on a particular topic. The study also demonstrates the potential for predicting users' choice of information sources based on their needs and preferences. These findings lay the foundation for developing personalized search, recommender, or intelligent systems that recommend relevant information sources based on user preferences. The study aims to contribute to information seeking research and provides a basis for further exploration. Additionally, future research could explore the influence of demographic and personal data on information source selection.

Similarly, Study 2 also has limitations. Using simulated task scenarios in a controlled laboratory setting may differ from the participants' daily lives and actual needs, potentially influencing their task completion and perception of problems and help. The study design prevented participants from seeking external assistance, which they might have done in real-life situations. The selection of potential problems and helps from a pre-defined list may not fully represent users' actual needs, potentially leading to inaccurate reports. The controlled setting and time constraints may also affect participants' responses. Furthermore, some measures relied on self-reported data, introducing subjective biases.

Despite these limitations, the study's main contributions lie in the detailed procedure for identifying and observing information needs throughout the information search process. It also sheds light on understanding barriers (reported problems) encountered during automated information retrieval, and the subsequent search behavior changes to mitigate those problems. The study examines the relationship between behaviors, perceived problems, and help in independent moments within a session, acknowledging that search sessions consist of interconnected strategies.

Regarding unified task modeling, a limitation is that the study only focuses on four elements of task states and omits other influential aspects, such as the emotional state. Nonetheless, this research provides a promising direction for conceptually deconstructing and computationally modeling complex search tasks from a process-oriented perspective. Future research can explore additional dimensions and enrich the task state framework, considering user characteristics and contextual factors that impact task states at various levels.

In terms of leveraging task representations, the experiments conducted on medium to large datasets, which were primarily curated in user studies, may lack real-world users' data and the inherent messiness found in actual search log data. This discrepancy could potentially impact the performance and generalizability of the experiments.

Since the curated datasets used in user studies are carefully designed and controlled, they may not fully capture the complexity and diversity of real-world search behavior. Real users' data, obtained from search logs or other sources, often exhibit various biases, noise, and unpredictability that are inherent to user interactions with search systems. By relying solely on curated datasets, the experiments may not fully capture the challenges and nuances that arise in real-world scenarios. Furthermore, curated datasets may not encompass the full range of tasks, information needs, and user behaviors encountered in the wild. The controlled nature of user studies allows for specific hypotheses to be tested, but it may limit the ability to fully explore the breadth and depth of user interactions in real-world search settings. However, the experiments show the importance of task information in retrieval and recommendations.

11.3 Future Work

In addition to the implications and future research directions mentioned above, the research presented in this thesis has additional opportunities for further work.

11.3.1 Task-specific Search Support Interface for Task Completion

The insights gained from this thesis offer the opportunity to seamlessly integrate task support into existing search user experiences. This integration can be achieved by introducing task-specific features, contextual suggestions, and guided interactions within the current search interfaces. The successful integration of task-specific support relies on understanding the context of user behavior, queries, and tasks. By leveraging our task modeling approach, search systems can effectively inform and personalize search results based on the specific tasks users are engaged in.

In specific scenarios, creating new search experiences that are specifically designed for particular tasks or domains can deliver a more focused and streamlined user experience. A notable example of this is Microsoft Bing's Bing Chat [1] experience, where users can engage in interactive conversations with the search system to refine queries, explore options, and receive personalized recommendations. These innovative approaches are especially advantageous for complex tasks that require interactive and conversational support. To further enhance task support, the task state representation model can aid the system in identifying users' challenges and offer tailored assistance in the form of task-related tips or proactive guidance to improve task completion.

[1]https://www.microsoft.com/en-us/edge/features/bing-chat?form=MT00D8

11.3.2 Modeling Tasks beyond Search

The concept of tasks holds great promise in diverse fields and user interaction systems, extending beyond search to encompass a wide range of user-centric online services. As such, we plan to focus our future research endeavors on advancing our task state model to encompass general user tasks across different devices, applications, and systems. This development will facilitate more effective contextual personalization for users, empowering system designers to target users based on their specific tasks and deliver tailored experiences. The valuable insights gained from this thesis can serve as a catalyst for further exploration in developing comprehensive and adaptable task models that extend beyond the confines of web search. Furthermore, our generic task extraction approach has potential applicability in various domains, opening up avenues for experimenting with broader task taxonomies, such as Anderson and Krathwohl–Bloom's taxonomy [5], to gain a deeper understanding of users' tasks in a general context.

11.3.3 Task Automation

In addition to task support and task completion, task autocompletion has emerged as another potential research area. Task automation interventions can be crucial in enhancing efficiency and streamlining processes. By automating repetitive tasks like data entry, note-taking, email generation, meeting scheduling, and even simple fact-finding search processes, systems can significantly reduce time and cognitive load for users. Leveraging techniques such as ACT-1 [2] or physics-informed AI [129], along with generative AI models, which harness natural language understanding and generation capabilities, can enable the automation of repetitive tasks and workflows. These approaches can facilitate task execution or provide guidance for performing specific actions. By integrating these techniques, interventions can assist users by providing relevant information, suggestions, or automated actions to support

[2]https://www.adept.ai/blog/act-1

task completion.

However, it is crucial to consider several factors that influence the applicability of interventions and task automation. These factors include the complexity of the tasks at hand, the availability of relevant data and resources, and the level of human oversight or decision-making required in the process. The feasibility and effectiveness of implementing interventions and task automation depend on the specific context and requirements of the tasks being addressed.

11.3.4 Ethical Considerations

As AI-powered systems become more advanced, there is a risk that they may influence user behavior in a biased way. This is particularly true for task-based intelligent systems, which may nudge users towards specific actions and perpetuate false trust. There is concern that these systems may create a feedback loop where users are no longer making decisions based on their own behavior but rather on the system's recommendations. This can lead to biases in the data and in users' decisions which can have unintended consequences. To address these issues, we need to consider the ethical implications of task modeling and ensure that we are not overtly influencing user behavior. We must also be aware of unintended consequences and ensure we are not creating biases in the data.

In conclusion, we encourage the active pursuit of this research direction, both in theoretical exploration and practical implementation. We envision that the task model and insights presented in this work can serve as a foundation for developing task-based intelligent assistance systems and advancing the field of interactive information retrieval. By integrating the insights and building blocks provided by this work, researchers and practitioners can contribute to developing more effective and personalized task-oriented support in various domains. We believe that further research and practical applications in this area will lead to significant advancements in the field and ultimately enhance the user experience in information seeking tasks.

REFERENCES

[1] Naresh Kumar Agarwal. Verifying survey items for construct validity: A two-stage sorting procedure for questionnaire design in information behavior research. *Proceedings of the ASIST Annual Meeting*, 48, 2011.

[2] Mohammad Aliannejadi, Morgan Harvey, Luca Costa, Matthew Pointon, and Fabio Crestani. Understanding mobile search task relevance and user behaviour in context. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, pages 143–151, 2019.

[3] Bryce Allen. Topic knowledge and online catalog search formulation. *The Library Quarterly*, 61(2):188–213, 1991.

[4] Claire J Anderson, Myron Glassman, R Bruce McAfee, and Thomas Pinelli. An investigation of factors affecting how engineers and scientists seek information. *Journal of Engineering and Technology Management*, 18(2):131–155, 2001.

[5] Lorin W Anderson, Benjamin Samuel Bloom, et al. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* Longman,, 2001.

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

[7] Susan J Ashford. Feedback-seeking in individual adaptation: A resource perspective. *Academy of Management journal*, 29(3):465–487, 1986.

[8] Kumaripaba Athukorala, Dorota Głowacka, Giulio Jacucci, Antti Oulasvirta, and Jilles Vreeken. Is exploratory search different? a comparison of information search behavior for exploratory and lookup tasks. *Journal of the Association for Information Science and Technology*, 67(11):2635–2651, 2016.

[9] Charles K Atkin. Anticipated communication and mass media information-seeking. *Public Opinion Quarterly*, 36(2):188–199, 1972.

[10] 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.

[11] Leif Azzopardi. Simulation of interaction: A tutorial on modelling and simulating user interaction and search behaviour. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 1227–1230, 2016.

[12] Leif Azzopardi, Maarten De Rijke, and Krisztian Balog. Building simulated queries for known-item topics: an analysis using six european languages. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 455–462, 2007.

[13] Leif Azzopardi, Kalervo Järvelin, Jaap Kamps, and Mark D Smucker. Simulation of interaction. 2010.

[14] Feza Baskaya, Heikki Keskustalo, and Kalervo Järvelin. Time drives interaction: Simulating sessions in diverse searching environments. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 105–114, 2012.

[15] Feza Baskaya, Heikki Keskustalo, and Kalervo Järvelin. Modeling behavioral factors ininteractive information retrieval. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pages 2297–2302, 2013.

[16] Marcia J Bates. An exploratory paradigm for online information retrieval. *Intelligent Information Systems for the Information Society. Amsterdam: North-Holland*, pages 91–99, 1986.

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

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

[19] N J Belkin. Anomalous states of knowledge as a basis for information retrieval, 1980.

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

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

[22] Nicholas J Belkin. A methodology for taking account of user tasks, goals and behavior for design of computerized library catalogs. *ACM SIGCHI Bulletin*, 23(1):61–65, 1991.

[23] Nicholas J Belkin. Salton award lecture: People, interacting with information. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1–2, 2015.

[24] Nicholas J Belkin. People, Interacting with Information. In *ACM SIGIR Forum*, volume 49, pages 13–27. ACM, 2016.

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

[26] 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.

[27] 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.

[28] 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.

[29] David J Bell and Ian Ruthven. Searcher's assessments of task complexity for web searching. In *European Conference on Information Retrieval*, pages 57–71. Springer, 2004.

[30] Gabriel Bénédict, Ruqing Zhang, and Donald Metzler. Gen-ir@ sigir 2023: The first workshop on generative information retrieval. *arXiv preprint arXiv:2306.02887*, 2023.

[31] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[32] Pia Borlund. The concept of relevance in ir. *Journal of the American Society for information Science and Technology*, 54(10):913–925, 2003.

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

[34] Horatiu Bota, Ke Zhou, and Joemon M Jose. Playing your cards right: The effect of entity cards on search behaviour and workload. In *Proceedings of the 2016 acm on conference on human information interaction and retrieval*, pages 131–140, 2016.

[35] Saskia Brand-Gruwel, Yvonne Kammerer, Ludo Van Meeuwen, and Tamara Van Gog. Source evaluation of domain experts and novices during web search. *Journal of Computer Assisted Learning*, 33(3):234–251, 2017.

[36] Kathy Brennan, Diane Kelly, and Jaime Arguello. The effect of cognitive abilities on information search for tasks of varying levels of complexity. In *Proceedings of the 5th Information Interaction in Context Symposium*, pages 165–174, 2014.

[37] Andrei Broder. A taxonomy of web search. *ACM SIGIR Forum*, 36(2):3, 2002.

[38] Bertram C Brookes. The developing cognitive viewpoint in information science. *De Mey, M. and al Editors*, pages 195–203, 1977.

[39] Michael K Buckland. Information as thing. *Journal of the American Society for Information Science (1986-1998)*, 42(5):351, 1991.

[40] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. Amazon's mechanical turk: A new source of inexpensive, yet high-quality data? 2016.

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

[42] Katriina Byström. Approaches to " task " in contemporary information studies. In *Proceedings of the Sixth International Conference on Conceptions of Library and Information Science—"Featuring the Future"*, volume 12, pages 1–10, 2007.

[43] Katriina Byström. Approaches to" task" in contemporary information studies. *Information research*, 12(4):12–4, 2007.

[44] 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.

[45] 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.

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

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

[48] Arthur Câmara, David Maxwell, and Claudia Hauff. Searching, learning, and subtopic ordering: A simulation-based analysis. In *European Conference on Information Retrieval*, pages 142–156. Springer, 2022.

[49] Donald J Campbell. Task complexity: A review and analysis. *Academy of management review*, 13(1):40–52, 1988.

[50] Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. Learning to rank: from pairwise approach to listwise approach. In *Proceedings of the 24th international conference on Machine learning*, pages 129–136, 2007.

[51] 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.

[52] 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.

[53] 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.

[54] Rafael Capurro and Birger Hjørland. The concept of information. *Annual review of information science and technology*, 37(1):343–411, 2003.

[55] Launor Carter, William Haythorn, and Margaret Howell. A further investigation of the criteria of leadership. *The Journal of Abnormal and Social Psychology*, 45(2):350, 1950.

[56] Ben Carterette, Ashraf Bah, and Mustafa Zengin. Dynamic test collections for retrieval evaluation. In *Proceedings of the 2015 international conference on the theory of information retrieval*, pages 91–100, 2015.

[57] Ben Carterette, Evangelos Kanoulas, Mark Hall, and Paul Clough. Overview of the trec 2014 session track. Technical report, Delaware Univ Newark Dept of Computer and Information Sciences, 2014.

[58] Ben Carterette, Evangelos Kanoulas, and Emine Yilmaz. Simulating simple user behavior for system effectiveness evaluation. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 611–620, 2011.

[59] Donald O Case and Lisa M Given. *Looking for information*. Emerald Group Publishing Limited, 2012.

[60] Donald O Case and Lisa M Given. *Looking for information*. Emerald Group Publishing Limited, 2012.

[61] Lara D Catledge and James E Pitkow. Characterizing browsing strategies in the world-wide web. *Computer Networks and ISDN systems*, 27(6):1065–1073, 1995.

[62] Jia Chen, Yiqun Liu, Cheng Luo, Jiaxin Mao, Min Zhang, and Shaoping Ma. Improving session search performance with a multi-mdp model. In *Asia Information Retrieval Symposium*, pages 45–59. Springer, 2018.

[63] 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.

[64] Justin Cheng, Jaime Teevan, Shamsi T Iqbal, and Michael S Bernstein. Break it down: A comparison of macro-and microtasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 4061–4064, 2015.

[65] E. Choi and C. Shah. Asking for more than an answer: What do askers expect in online Q&A services? *Journal of Information Science*, 42(2):0165551516645530, 2016.

[66] Erik Choi and Chirag Shah. User motivations for asking questions in online Q&A services. *Journal of the Association for Information Science and Technology*, 67(5):1182–1197, 2016.

[67] 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.

[68] Edward W Christensen and James R Bailey. A source accessibility effect on media selection. *Management Communication Quarterly*, 10(3):373–387, 1997.

[69] Aleksandr Chuklin, Ilya Markov, and Maarten de Rijke. Click models for web search. *Synthesis lectures on information concepts, retrieval, and services*, 7(3):1–115, 2015.

[70] Aleksandr Chuklin, Pavel Serdyukov, and Maarten De Rijke. Click model-based information retrieval metrics. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 493–502, 2013.

[71] Cyril Cleverdon. The cranfield tests on index language devices. In *Aslib proceedings*. MCB UP Ltd, 1967.

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

[73] Charles Cole. Information Need and the Beginning of Information Search. *Encyclopedia of Information Science and Technology, Third Edition*, (1968):4117–4128, 2015.

[74] Michael J Cole, Jacek Gwizdka, Chang Liu, Ralf Bierig, Nicholas J Belkin, and Xiangmin Zhang. Task and user effects on reading patterns in information search. *Interacting with Computers*, 23(4):346–362, 2011.

[75] 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.

[76] Robert Cooley, Bamshad Mobasher, and Jaideep Srivastava. Data preparation for mining world wide web browsing patterns. *Knowledge and information systems*, 1(1):5–32, 1999.

[77] Gordon V Cormack and Thomas R Lynam. Statistical precision of information retrieval evaluation. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 533–540, 2006.

[78] Ian Cornelius. Theorizing information for information science. *Annual review of information science and technology*, 36(1):392–425, 2002.

[79] John W Creswell and J Creswell. *Research design*. Sage publications Thousand Oaks, CA, 2003.

[80] 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.

[81] Zhuyun Dai and Jamie Callan. Deeper text understanding for ir with contextual neural language modeling. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*, pages 985–988, 2019.

[82] Brenda Dervin. *An overview of sense-making research: Concepts, methods, and results to date*. School of Communications, University of Washington, Seattle, 1983.

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

[84] Brenda Dervin, Lois Foreman-Wernet, and Eric Lauterbach. *Sense-making methodology reader: Selected writings of Brenda Dervin*. Hampton Press, Cresskill, NJ, 2003.

[85] Brenda Dervin and Michael Nilan. *Information needs and uses*, volume 21. 1986.

[86] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[87] Laura Dietz, Bhaskar Mitra, Jeremy Pickens, Hana Anber, Sandeep Avula, Asia Biega, Adrian Boteanu, Shubham Chatterjee, Jeff Dalton, Laura Dietz, et al. Report on the first hipstir workshop on the future of information retrieval. In *ACM Sigir forum*, volume 53, pages 62–75. ACM New York, NY, USA, 2021.

[88] Georges E Dupret and Benjamin Piwowarski. A user browsing model to predict search engine click data from past observations. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 331–338, 2008.

[89] Carsten Eickhoff, Jaime Teevan, Ryen White, and Susan Dumais. Lessons from the journey: a query log analysis of within-session learning. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 223–232, 2014.

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

[91] Shaohua Fan, Junxiong Zhu, Xiaotian Han, Chuan Shi, Linmei Hu, Biyu Ma, and Yongliang Li. Metapath-guided heterogeneous graph neural network for intent recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2478–2486, 2019.

[92] 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.

[93] Karen E Fisher and Charles M Naumer. Information grounds: Theoretical basis and empirical findings on information flow in social settings. *New directions in human information behavior*, pages 93–111, 2006.

[94] Luciano Floridi. Ai as agency without intelligence: On chatgpt, large language models, and other generative models. *Philosophy & Technology*, 36(1):15, 2023.

[95] Steve Fox, Kuldeep Karnawat, Mark Mydland, Susan Dumais, and Thomas White. Evaluating implicit measures to improve web search. *ACM Transactions on Information Systems (TOIS*, 23(2), 2005.

[96] Xinyu Fu, Jiani Zhang, Ziqiao Meng, and Irwin King. Magnn: Metapath aggregated graph neural network for heterogeneous graph embedding. In *Proceedings of The Web Conference 2020*, pages 2331–2341, 2020.

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

[98] Maarten Grootendorst. Keybert: Minimal keyword extraction with bert. *Internet]. Available: https://maartengr. github. io/KeyBERT/index. html*, 2020.

[99] Jacek Gwizdka, Preben Hansen, Claudia Hauff, Jiyin He, and Noriko Kando. Search as learning (sal) workshop 2016. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 1249–1250, 2016.

[100] Jacek Gwizdka and Ian Spence. What can searching behavior tell us about the difficulty of information tasks? a study of web navigation. *Proceedings of the American Society for Information Science and Technology*, 43(1):1–22, 2006.

[101] J Richard Hackman. Effects of task characteristics on group products. *Journal of Experimental Social Psychology*, 4(2):162–187, 1968.

[102] J Richard Hackman and Charles G Morris. Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. *Advances in experimental social psychology*, 8:45–99, 1975.

[103] JoAnn T Hackos and Janice Redish. *User and task analysis for interface design*, volume 1. Wiley New York, 1998.

[104] Matthias Hagen, Jakob Gomoll, Anna Beyer, and Benno Stein. From search session detection to search mission detection. In *Proceedings of the 10th Conference on Open Research Areas in Information Retrieval*, pages 85–92, 2013.

[105] Ahmed Hassan Awadallah, Cathal Gurrin, Mark Sanderson, and Ryen W White. Task intelligence workshop@ wsdm 2019. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 848–849, 2019.

[106] Ahmed Hassan Awadallah, Ryen W White, Patrick Pantel, Susan T Dumais, and Yi-Min Wang. Supporting complex search tasks. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*, pages 829–838, 2014.

[107] Helene A Hembrooke, Laura A Granka, Geraldine K Gay, and Elizabeth D Liddy. The effects of expertise and feedback on search term selection and subsequent learning. *Journal of the American Society for Information Science and Technology*, 56(8):861–871, 2005.

[108] 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.

[109] Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. Session-based recommendations with recurrent neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, 2015.

[110] Daniel Hienert, Matthew Mitsui, Philipp Mayr, Chirag Shah, and Nicholas J. Belkin. The Role of the Task Topic in Web Search of Different Task Types. *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - CHIIR '18*, pages 72–81, 2018.

[111] Ingrid Hsieh-Yee. Effects of search experience and subject knowledge on the search tactics of novice and experienced searchers. *Journal of the american society for information science*, 44(3):161–174, 1993.

[112] Rong Hu, Kun Lu, and Soohyung Joo. Effects of topic familiarity and search skills on query reformulation behavior. *Proceedings of the American Society for Information Science and Technology*, 50(1):1–9, 2013.

[113] P Ingwersen. Järvelin. k.(2005b). the turn: Integration of information seeking and retrieval in context, 2005.

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

[115] Peter Ingwersen and Kalervo Järvelin. *The turn: Integration of information seeking and retrieval in context*, volume 18. Springer Science & Business Media, 2006.

[116] Emi Ishita, Yosuke Miyata, Shuichi Ueda, and Keiko Kurata. A structural equation model of information retrieval skills. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, pages 317–320, 2017.

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

[118] Bernard J Jansen, Amanda Spink, and Vinish Kathuria. How to define searching sessions on web search engines. In *International Workshop on Knowledge Discovery on the Web*, pages 92–109. Springer, 2006.

[119] Bernard J Jansen, Amanda Spink, and Tefko Saracevic. Real life, real users, and real needs: a study and analysis of user queries on the web. *Information processing & management*, 36(2):207–227, 2000.

[120] Kalervo Järvelin and Peter Ingwersen. Information seeking research needs extension towards tasks and technology. *Information Research: An International Electronic Journal*, 10(1):n1, 2004.

[121] 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.

[122] 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, 2014.

[123] 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.

[124] Rosie Jones and Kristina Lisa Klinkner. Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. In *Proceedings of the 17th ACM conference on Information and knowledge manage- ment*, pages 699–708. ACM, 2008.

[125] Rosie Jones and Kristina Lisa Klinkner. Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 699–708, 2008.

[126] Rosie Jones, Benjamin Rey, Omid Madani, and Wiley Greiner. Generating query substitutions. In *Proceedings of the 15th international conference on World Wide Web*, pages 387–396, 2006.

[127] Kalervo Järvelin. *Evaluation*, page 113–138. Facet, 2011.

[128] Jarkko Kari. Making sense of sense-making: From metatheory to substantive theory in the context of paranormal information seeking. In *Nordis-Net workshop:(Meta) theoretical stands in studying library and information institutions: individual, organizational and societal aspects, Oslo, Norway*, pages 1–20. Citeseer, 1998.

[129] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

[130] Atoosa Kasirzadeh and Iason Gabriel. In conversation with artificial intelligence: aligning language models with human values. *Philosophy & Technology*, 36(2):1–24, 2023.

[131] Elihu Katz, Hadassah Haas, and Michael Gurevitch. On the use of the mass media for important things. *American sociological review*, pages 164–181, 1973.

[132] Elihu Katz, Hadassah Haas, and Michael Gurevitch. On the use of the mass media for important things. *American sociological review*, pages 164–181, 1973.

[133] Leonard Kaufman and Peter J Rousseeuw. Partitioning around medoids (program pam). *Finding groups in data: an introduction to cluster analysis*, 344:68–125, 1990.

[134] 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.

[135] 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.

[136] Diane Kelly. Measuring online information seeking context, part 1: Background and method. *Journal of the American Society for Information Science and Technology*, 57(13):1729–1739, 2006.

[137] Diane Kelly. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends in Information Retrieval*, 3(12), 2009.

[138] 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.

[139] 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.

[140] Diane Kelly and Colleen Cool. The effects of topic familiarity on information search behavior. In *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, pages 74–75, 2002.

[141] Jeonghyun Kim. Task as a predictable indicator for information seeking behavior on the web, 2006.

[142] Jeonghyun Kim. Modeling task-based information seeking on the web: Application of information seeking strategy schema. *Proceedings of the American Society for Information Science and Technology*, 44(1):1–13, 2008.

[143] Alexander Kotov, Paul N Bennett, Ryen W White, Susan T Dumais, and Jaime Teevan. Modeling and analysis of cross-session search tasks. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 5–14, 2011.

[144] Michael Kotzyba, Tatiana Gossen, Johannes Schwerdt, and Andreas Nürnberger. Exploration or fact-finding: Inferring user's search activity just in time. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*, pages 87–96, 2017.

[145] C Kuhlthau and Pertii Vakkari. Information Seeking in Context (ISIC). *Information Processing & Management*, 35(1999):723–725, 1999.

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

[147] C. C. Kuhlthau. Inside the search process: seeking meaning from the users perspective. *Journal of the American Society for Information Science*, 42(5):361–371, 1991.

[148] Carol Collier Kuhlthau. *Seeking meaning: A process approach to library and information services*. Libraries Unltd Incorporated, 2004.

[149] Sanna Kumpulainen and Kalervo Järvelin. Information interaction in molecular medicine: integrated use of multiple channels. In *Proceedings of the third symposium on Information interaction in context*, pages 95–104, 2010.

[150] 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.

[151] Daniel Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, 2000.

[152] Liangda Li, Hongbo Deng, Anlei Dong, Yi Chang, and Hongyuan Zha. Identifying and labeling search tasks via query-based hawkes processes. In *Proceedings of the 20th*

ACM SIGKDD international conference on Knowledge discovery and data mining, pages 731–740, 2014.

[153] Ruirui Li, Ben Kao, Bin Bi, Reynold Cheng, and Eric Lo. Dqr: A probabilistic approach to diversified query recommendation. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 16–25, 2012.

[154] 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.

[155] Yuelin Li. An exploration of the relationships between work tasks and users' interaction performance. *Proceedings of the ASIST Annual Meeting*, 47, 2010.

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

[157] 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.

[158] Yang Lin, Charles Cole, and Kimiz Dalkir. The relationship between perceived value and information source use during KM strategic decision-making: A study of 17 Chinese business managers. *Information Processing and Management*, 50(1):156–174, 2014.

[159] 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.

[160] 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.

[161] Chang Liu, Jingjing Liu, Michael Cole, Nicholas J Belkin, and Xiangmin Zhang. Task difficulty and domain knowledge effects on information search behaviors. *Proceedings of the American Society for Information Science and Technology*, 49(1):1–10, 2012.

[162] Chang Liu and Yiming Wei. The impacts of time constraint on users' search strategy during search process. *Proceedings of the Association for Information Science and Technology*, 53(1):1–9, 2016.

[163] Chang Liu, Xiangmin Zhang, and Wei Huang. The exploration of objective task difficulty and domain knowledge effects on users' query formulation. *Proceedings of the Association for Information Science and Technology*, 53(1):1–9, 2016.

[164] Jingjing Liu and Nicholas J Belkin. Personalizing information retrieval for multi-session tasks: The roles of task stage and task type. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 26–33. ACM, 2010.

[165] Jingjing Liu and Nicholas J. Belkin. Personalizing information retrieval for multi-session tasks: The roles of task stage and task type. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 26–33. ACM, 2010.

[166] 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, 2010.

[167] 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.

[168] Jingjing Liu, Chang Liu, and Nicholas J Belkin. Predicting information searchers' topic knowledge at different search stages. *Journal of the Association for Information Science and Technology*, 67(11):2652–2666, 2016.

[169] Jingjing Liu, Chang Liu, Jacek Gwizdka, and Nicholas J Belkin. Can search systems detect users' task difficulty? some behavioral signals. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 845–846, 2010.

[170] Jiqun Liu. *A State-based Approach to Supporting Users in Complex Search Tasks*. PhD thesis, Rutgers The State University of New Jersey, School of Graduate Studies, 2020.

[171] 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.

[172] 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.

[173] Jiqun Liu, Shawon Sarkar, and Chirag Shah. Identifying and predicting the states of complex search tasks. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, pages 193–202, 2020.

[174] 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.

[175] Jiqun Liu and Ran Yu. State-aware meta-evaluation of evaluation metrics in interactive information retrieval. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, CIKM '21, page 3258–3262, New York, NY, USA, 2021. Association for Computing Machinery.

[176] Robert M. Losee. A discipline independent definition of information. *Journal of the American Society for Information Science*, 48(3):254–269, 1997.

[177] Matt Lovett, Saleh Bajaba, Myra Lovett, and Marcia J Simmering. Data quality from crowdsourced surveys: A mixed method inquiry into perceptions of amazon's mechanical turk masters. *Applied Psychology*, 67(2):339–366, 2018.

[178] Claudio Lucchese, Salvatore Orlando, Raffaele Perego, Fabrizio Silvestri, and Gabriele Tolomei. Detecting task-based query sessions using collaborative knowledge. In *Web Intelligence and Intelligent Agent Technology (WIIAT), 2010 IEEE/WIC/ACM International Conference on*, volume 3, pages 128–131. IEEE, 2010.

[179] Claudio Lucchese, Salvatore Orlando, Raffaele Perego, Fabrizio Silvestri, and Gabriele Tolomei. Identifying task-based sessions in search engine query logs. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 277–286. ACM, 2011.

[180] Claudio Lucchese, Salvatore Orlando, Raffaele Perego, Fabrizio Silvestri, and Gabriele Tolomei. Discovering tasks from search engine query logs. *ACM Transactions on Information Systems (TOIS*, 31(3), 2013.

[181] 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.

[182] Andrew D Madden. Evolution and information. *Journal of Documentation*, 60(1):9–23, 2004.

[183] ML Maehr. Meaning and motivation: toward a theory of personal investment. en r. ames and c. ames (eds.): Research on motivation in education: vol. 1. student motivation (p. 144), 1984.

[184] Stepan Malkevich, Ilya Markov, Elena Michailova, and Maarten De Rijke. Evaluating and analyzing click simulation in web search. In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*, pages 281–284, 2017.

[185] Christopher D Manning and Prabhakar Raghavan. and schutze, h.[2008] introduction to information retrieval, 2008.

[186] Kelong Mao, Zhicheng Dou, Haonan Chen, Fengran Mo, and Hongjin Qian. Large language models know your contextual search intent: A prompting framework for conversational search. *arXiv preprint arXiv:2303.06573*, 2023.

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

[188] Gary Marchionini. *Information seeking in electronic environments.* Number 9. Cambridge university press, 1997.

[189] Gary Marchionini. *Information seeking in electronic environments.* Number 9. Cambridge university press, 1997.

[190] Gary Marchionini. Information concepts: From books to cyberspace identities. *Synthesis Lectures on Information Concepts, Retrieval, and Services*, 2(1):1–105, 2010.

[191] Gary Marchionini and Hermann Maurer. The roles of digital libraries in teaching and learning. *Communications of the ACM*, 38(4):67–75, 1995.

[192] David Maxwell and Leif Azzopardi. Agents, simulated users and humans: An analysis of performance and behaviour. In *Proceedings of the 25th ACM international on conference on information and knowledge management*, pages 731–740, 2016.

[193] David Maxwell and Leif Azzopardi. Simulating interactive information retrieval: Simiir: A framework for the simulation of interaction. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 1141–1144, 2016.

[194] David Maxwell, Leif Azzopardi, Kalervo Järvelin, and Heikki Keskustalo. Searching and stopping: An analysis of stopping rules and strategies. In *Proceedings of the 24th ACM international on conference on information and knowledge management*, pages 313–322, 2015.

[195] Maureen McCreadie and Ronald E Rice. Trends in analyzing access to information. Part I: cross-disciplinary conceptualizations of access. *Information processing & management*, 35(1):45–76, 1999.

[196] Aaron F McDaid, Derek Greene, and Neil Hurley. Normalized mutual information to evaluate overlapping community finding algorithms. *arXiv preprint arXiv:1110.2515*, 2011.

[197] Molly McGregor, Leif Azzopardi, and Martin Halvey. Untangling cost, effort, and load in information seeking and retrieval. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*, pages 151–161, 2021.

[198] Rishabh Mehrotra. *Inferring user needs and tasks from user interactions*. PhD thesis, UCL (University College London), 2018.

[199] Rishabh Mehrotra, Prasanta Bhattacharya, and Emine Yilmaz. Deconstructing complex search tasks: a bayesian nonparametric approach for extracting sub-tasks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 599–605,, 2016.

[200] 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.

[201] Rishabh Mehrotra and Emine Yilmaz. Task embeddings: Learning query embeddings using task context. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 2199–2202. ACM, 2017.

[202] Donald Metzler, Yi Tay, Dara Bahri, and Marc Najork. Rethinking search: making domain experts out of dilettantes. In *ACM SIGIR Forum*, volume 55, pages 1–27. ACM New York, NY, USA, 2021.

[203] 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.

[204] 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.

[205] Matthew Mitsui, Jiqun Liu, and Chirag Shah. How much is too much? whole session vs. first query behaviors in task type prediction. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 1141–1144, 2018.

[206] Matthew Mitsui, Jiqun Liu, and Chirag Shah. The paradox of personalization: Does task prediction require individualized models? In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, pages 277–280, 2018.

[207] Matthew Mitsui and Chirag Shah. Bridging Gaps: Predicting User and Task Characteristics from Partial User Information. In *Proceedings of the 42Nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR'19, pages 415–424, New York, NY, USA, 2019. ACM.

[208] Matthew Mitsui and Chirag Shah. Bridging gaps: Predicting user and task characteristics from partial user information. In *Proceedings of the 42Nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR'19, pages 415–424, New York, NY, USA, 2019. ACM.

[209] Matthew Mitsui and Chirag Shah. Bridging gaps: Predicting user and task characteristics from partial user information. In *Proceedings of the 42nd International ACM*

SIGIR Conference on Research and Development in Information Retrieval, pages 415–424, 2019.

[210] 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.

[211] 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.

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

[213] Don Norman. *The design of everyday things: Revised and expanded edition.* Basic books, 2013.

[214] Sanghee Oh. The characteristics and motivations of health answerers for sharing information, knowledge, and experiences in online environments. *Journal of the Association for Information Science and Technology*, 63(3):543–557, 2012.

[215] Sanghee Oh. The characteristics and motivations of health answerers for sharing information, knowledge, and experiences in online environments. *Journal of the Association for Information Science and Technology*, 63(3):543–557, 2012.

[216] José Luis Ortega and Isidro Aguillo. Differences between web sessions according to the origin of their visits. *Journal of Informetrics*, 4(3):331–337, 2010.

[217] Fabio Paternò. Concurtasktrees: an engineered notation for task models. *The handbook of task analysis for human-computer interaction*, pages 483–503, 2004.

[218] Allan D Pratt. The information of the image. *Libri*, 27(1):204–220, 1977.

[219] Tao Qin and Tie-Yan Liu. Introducing letor 4.0 datasets. *arXiv preprint arXiv:1306.2597*, 2013.

[220] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

[221] Filip Radlinski and Nick Craswell. Comparing the sensitivity of information retrieval metrics. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 667–674, 2010.

[222] Filip Radlinski and Thorsten Joachims. Query chains: learning to rank from implicit feedback. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 239–248, 2005.

[223] Filip Radlinski, Martin Szummer, and Nick Craswell. Inferring query intent from reformulations and clicks. In *Proceedings of the 19th international conference on World wide web*, pages 1171–1172, 2010.

[224] Jane Reid. A task-oriented non-interactive evaluation methodology for information retrieval systems. *Information Retrieval*, 2(1):115–129, 2000.

[225] CarrieLynn D Reinhard and Brenda Dervin. Comparing situated sense-making processes in virtual worlds: Application of dervin's sense-making methodology to media reception situations. *Convergence*, 18(1):27–48, 2012.

[226] Eun Youp Rha, Matthew Mitsui, Nicholas J Belkin, and Chirag Shah. Exploring the relationships between search intentions and query reformulations. In *Proceedings of the 79th ASIS&T Annual Meeting: Creating Knowledge, Enhancing Lives through Information & Technology*, page 48. American Society for Information Science, 2016.

[227] 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.

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

[229] 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.

[230] Ryan A Rossi, Nesreen K Ahmed, Aldo Carranza, David Arbour, Anup Rao, Sungchul Kim, and Eunyee Koh. Heterogeneous graphlets. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 15(1):1–43, 2020.

[231] Brent D Ruben. General system theory: An approach to human communication. *Approaches to human communication*, pages 120–144, 1972.

[232] Diane Liang Rulke, Srilata Zaheer, and Marc H Anderson. Sources of managers' knowledge of organizational capabilities. *Organizational behavior and human decision processes*, 82(1):134–149, 2000.

[233] Daniel M Russell. *The Joy of Search: A Google Insider's Guide to Going Beyond the Basics*. MIT Press, 2019.

[234] Ian Ruthven. Interactive information retrieval. *Annual review of information science and technology*, 42(1):43–91, 2008.

[235] Miamaria Saastamoinen and Kalervo Järvelin. Search task features in work tasks of varying types and complexity. *Journal of the Association for Information Science and Technology*, 68(5):1111–1123, 2017.

[236] Miamaria Saastamoinen and Sanna Kumpulainen. Expected and materialised information source use by municipal officials: intertwining with task complexity. 2014.

[237] Mark Sanderson and W Bruce Croft. The history of information retrieval research. *Proceedings of the IEEE*, 100(Special Centennial Issue):1444–1451, 2012.

[238] Tefko Saracevic. Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the Association for Information Science and Technology*, 26(6):321–343, 1975.

[239] Tefko Saracevic. Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the Association for Information Science and Technology*, 26(6):321–343, 1975.

[240] Tefko Saracevic and Paul Kantor. A study of information seeking and retrieving. iii. searchers, searches, and overlap. *Journal of the American Society for information Science*, 39(3):197–216, 1988.

[241] Shawon Sarka, Yiwei Wang, and Chirag Shah. Investigating relations of information seeking outcomes to the selection and use of information sources. *Proceedings of the Association for Information Science and Technology*, 54(1):347–356, 2017.

[242] Shawon Sarkar, Matthew Mitsui, Jiqun Liu, and Chirag Shah. Implicit information need as explicit problems, help, and behavioral signals. *Information Processing & Management*, 57(2):102069, 2020.

[243] Shawon Sarkar and Chirag Shah. An integrated model of task, information needs, sources and uncertainty to design task-aware search systems. In *Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval*, pages 83–92, 2021.

[244] Shawon Sarkar and Chirag Shah. A synthetic search session generator for task-aware information seeking and retrieval. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, WSDM '23, page 1156–1159, New York, NY, USA, 2023. Association for Computing Machinery.

[245] 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.

[246] 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.

[247] Reijo Savolainen. Source preferences in the context of seeking problem-specific information. *Information Processing & Management*, 44(1):274–293, 2008.

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

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

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

[251] Reijo Savolainen. Elaborating the conceptual space of information-seeking phenomena. *Information Research: An International Electronic Journal*, 21(3):n3, 2016.

[252] 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, 2018.

[253] Chirag Shah and Emily M Bender. Situating search. In *ACM SIGIR Conference on Human Information Interaction and Retrieval*, pages 221–232, 2022.

[254] 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.

[255] Chirag Shah and Vanessa Kitzie. Social Q&A and virtual reference - comparing apples and oranges with the help of experts and users. *Journal of the American Society for Information Science and Technology*, 63(10):2020–2036, 2012.

[256] Chirag Shah, Vanessa Kitzie, and Erik Choi. Modalities, motivations, and materials–investigating traditional and social online q&a services. *Journal of Information Science*, 40(5):669–687, 2014.

[257] Chirag Shah and Jennifer Sonne. Seeking information in online environments - where, who, and why? *iConference 2015 Proceedings*, 2015.

[258] Chirag Shah, Ryen White, Paul Thomas, Bhaskar Mitra, Shawon Sarkar, and Nicholas Belkin. Taking search to task. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval*, pages 1–13, 2023.

[259] Chirag Shah and Ryen W White. Tutorial on task-based search and assistance. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2436–2439, 2020.

[260] Chirag Shah and Ryen W White. Task intelligence for search and recommendation. *Synthesis Lectures onSynthesis Lectures on Information Concepts, Retrieval, and Services*, 13(3):1–160, 2021.

[261] Claude E Shannon and Warren Weaver. *The mathematical theory of information*. University of Illinois press, 1949.

[262] Lee S Shulman. Those who understand: Knowledge growth in teaching. *Educational researcher*, 15(2):4–14, 1986.

[263] Myra Spiliopoulou, Bamshad Mobasher, Bettina Berendt, and Miki Nakagawa. A framework for the evaluation of session reconstruction heuristics in web-usage analysis. *Informs journal on computing*, 15(2):171–190, 2003.

[264] Jian-tao Sun, Dou Shen, Qiang Yang, and Zheng Chen. Building bridges for web query classification, August 10 2010. US Patent 7,774,360.

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

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

[267] Robert S Taylor. Question-negotiation and information seeking in libraries. *College & research libraries*, 29(3):178–194, 1968.

[268] Jaime Teevan, Susan T Dumais, and Daniel J Liebling. To personalize or not to personalize: modeling queries with variation in user intent. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 163–170, 2008.

[269] Gabriele Tolomei, Salvatore Orlando, and Fabrizio Silvestri. Towards task-based search and recommender systems. In *Proceedings of the IEEE International Conference on Data Engineering Workshops*, pages 333–336, 2010.

[270] Elaine G. Toms. *Task-based information searching and retrieval*, page 43–60. Facet, 2011.

[271] Elaine G Toms, Heather O'Brien, Tayze Mackenzie, Chris Jordan, Luanne Freund, Sandra Toze, Emilie Dawe, and Alexandra Macnutt. Task effects on interactive search: The query factor. In *International Workshop of the Initiative for the Evaluation of XML Retrieval*, pages 359–372. Springer, 2007.

[272] Cunchao Tu, Han Liu, Zhiyuan Liu, and Maosong Sun. Cane: Context-aware network embedding for relation modeling. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1722–1731, 2017.

[273] Hossein Vahabi, Margareta Ackerman, David Loker, Ricardo Baeza-Yates, and Alejandro Lopez-Ortiz. Orthogonal query recommendation. In *Proceedings of the 7th ACM Conference on Recommender Systems*, pages 33–40, 2013.

[274] Pertti Vakkari. Task complexity, problem structure and information actions: Integrating studies on information seeking and retrieval. *Information processing & management*, 35(6):819–837, 1999.

[275] Pertti Vakkari. Relevance and contributing information types of searched documents in task performance. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 2–9, 2000.

[276] Pertti Vakkari. A theory of the task-based information retrieval process: A summary and generalisation of a longitudinal study. *Journal of documentation*, 2001.

[277] Pertti Vakkari. Task-based information searching. *Annual review of information science and technology*, 37(1):413–464, 2003.

[278] Pertti Vakkari and Nanna Hakala. Changes in relevance criteria and problem stages in task performance. *Journal of documentation*, 2000.

[279] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[280] Suzan Verberne, Maya Sappelli, Kalervo Järvelin, and Wessel Kraaij. User simulations for interactive search: Evaluating personalized query suggestion. In *European Conference on Information Retrieval*, pages 678–690. Springer, 2015.

[281] Ellen M Voorhees. The philosophy of information retrieval evaluation. In *Workshop of the cross-language evaluation forum for european languages*, pages 355–370. Springer, 2001.

[282] Chong Wang, Achir Kalra, Cristian Borcea, and Yi Chen. Webpage depth-level dwell time prediction. In *Proceedings of the 25th acm international on conference on information and knowledge management*, pages 1937–1940, 2016.

[283] Hongning Wang, Yang Song, Ming-Wei Chang, Xiaodong He, Ryen W White, and Wei Chu. Learning to extract cross-session search tasks. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1353–1364, 2013.

[284] RuiXing Wang, Kuan Fang, RiKang Zhou, Zhan Shen, and LiWen Fan. Serank: Optimize sequencewise learning to rank using squeeze-and-excitation network. *arXiv preprint arXiv:2006.04084*, 2020.

[285] Yiwei Wang, Shawon Sarkar, and Chirag Shah. Investigating information seekers' selection of interpersonal and impersonal sources. In *Proceedings of the 2017 conference on conference human information interaction and retrieval*, pages 353–356, 2017.

[286] Yiwei Wang, Shawon Sarkar, and Chirag Shah. Juggling with information sources, task type, and information quality. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*, pages 82–91, 2018.

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

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

[289] Ryen W White, Susan T Dumais, and Jaime Teevan. Characterizing the influence of domain expertise on web search behavior. In *Proceedings of the second ACM international conference on web search and data mining*, pages 132–141, 2009.

[290] Ryen W White and Jeff Huang. Assessing the scenic route: measuring the value of search trails in web logs. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 587–594. ACM, 2010.

[291] Ryen W White, Elnaz Nouri, James Woffinden-Luey, Mark EncarnacióN, and Sujay Kumar Jauhar. Microtask detection. *ACM Transactions on Information Systems (TOIS)*, 39(2):1–29, 2021.

[292] Ryen W White and Resa A Roth. Exploratory search: Beyond the query-response paradigm. *Synthesis lectures on information concepts, retrieval, and services*, 1(1):1–98, 2009.

[293] Barbara Wildemuth, Luanne Freund, and Elaine G. Toms. Untangling search task complexity and difficulty in the context of interactive information retrieval studies. *Journal of Documentation*, 70(6):1118–1140, 2014.

[294] 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.

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

[296] Thomas Daniel Wilson. On User Studies And Information Needs. *Journal Of Documentation*, 37(1):3 – 15, 1981.

[297] Tom D Wilson. On user studies and information needs. *Journal of documentation*, 1981.

[298] I-Chin Wu, Duen-Ren Liu, and Pei-Cheng Chang. Toward incorporating a task-stage identification technique into the long-term document support process. *Information processing & management*, 44(5):1649–1672, 2008.

[299] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. Session-based recommendation with graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 346–353, 2019.

[300] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. Session-based recommendation with graph neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 346–353, 2019.

[301] 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.

[302] Wan-Ching Wu, Diane Kelly, and Avneesh Sud. Using Information Scent and Need for Cognition to Understand Online Search Behavior. In *Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval*, volume 37, pages 557–566, 2014.

[303] Hong Xie. Planned and situated aspects in interactive ir: Patterns of user interactive intentions and information seeking strategies. In *Proceedings of the ASIST Annual Meeting*, volume 34, pages 101–10, 1997.

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

[305] 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.

[306] Qianli Xing, Yiqun Liu, Jian-Yun Nie, Min Zhang, Shaoping Ma, and Kuo Zhang. Incorporating user preferences into click models. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pages 1301–1310, 2013.

[307] Yunjie Xu, Cheng Yian Tan, and Li Yang. Who will you ask? An empirical study of interpersonal task information seeking. *Journal of the American Society for Information Science and Technology*, 57(12):1666–1677, 2006.

[308] Ka Yee Yeung and Walter L Ruzzo. Details of the adjusted rand index and clustering algorithms, supplement to the paper an empirical study on principal component analysis for clustering gene expression data. *Bioinformatics*, 17(9):763–774, 2001.

[309] Jeffrey M Zacks, Barbara Tversky, and Gowri Iyer. Perceiving, remembering, and communicating structure in events. *Journal of experimental psychology: General*, 130(1):29, 2001.

[310] Mustafa Zengin. *User Simulations in Search Sessions*. PhD thesis, University of Delaware, 2019.

[311] Xiangmin Zhang, Jingjing Liu, and Michael Cole. Task topic knowledge vs. background domain knowledge: Impact of two types of knowledge on user search performance. In *Advances in information systems and technologies*, pages 179–191. Springer, 2013.

[312] Yan Zhang. Contextualizing consumer health information searching: an analysis of questions in a social q&a community. In *Proceedings of the 1st ACM International Health Informatics Symposium*, pages 210–219. ACM, 2010.

[313] Yan Zhang. Contextualizing consumer health information searching: an analysis of questions in a social Q&A community. In *Proceedings of the 1st ACM International Health Informatics Symposium*, pages 210–219. ACM, 2010.

[314] Yao Zhang and Chang Liu. Users' knowledge use and change during information searching process: A perspective of vocabulary usage. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, pages 47–56, 2020.

[315] Xiaofeng Zhu and Diego Klabjan. Listwise learning to rank by exploring unique ratings. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 798–806, 2020.