Task, Information Seeking Intentions, and User Behavior: Toward A Multi-level Understanding of Web Search

Jiqun Liu, Matthew Mitsui, Nicholas J. Belkin, Chirag Shah
School of Communication & Information
Rutgers University
New Brunswick, NJ, USA, 08901
{jl2033,mmitsui,belkin,chirags}@scarletmail.rutgers.edu

ABSTRACT

According to the cognitive viewpoint of information retrieval (IR) research, a search task can be conceptualized as a sequence of information seeking intentions which both motivate and are influenced by search behaviors. While the behavioral effects of task features have been thoroughly discussed in a large body of literature, how different information seeking intentions in query segments serve as bridges between task and Web search behavior still remains unexplored. To develop a more comprehensive, multi-level (i.e., task level, intention level, and behavior level) understanding of Web search, the authors analyzed intention and search behavior data collected from 693 query segments generated by 40 participants in a controlled lab setting, seeking to answer two main research questions: 1) *from task to intention*: how do different task features affect users' information seeking intentions at different stages of a search session? 2) *from intention to behavior*: How is a user's search behavior associated with their information seeking intentions in the current and next query segments respectively? The results demonstrate that: 1) Task features significantly affected the frequency of occurrence of most of the information seeking intentions, and these effects gradually faded away as search sessions proceeded; 2) The presences of a variety of intentions in both current and subsequent query segments were connected with and detectable by different subsets of behavioral measures. This study contributes to the understanding of the connections between task, intentions in query segments, and search behavior, and thereby has implications for designing system affordances for supporting different intentions and search activities in various task stages and contexts.

CCS CONCEPTS

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

KEYWORDS

Information seeking intention; task; interactive IR

CHIIR '19, March 10–14, 2019, Glasgow, United Kingdom

ACM ISBN 978-1-4503-6025-8/19/03…$15.00
https://doi.org/10.1145/3295750.3298922

ACM Reference Format:
Jiqun Liu, Matthew Mitsui, Nicholas J. Belkin, Chirag Shah. 2019. Task, Information Seeking Intentions, and User Behavior: Toward A Multi-level Understanding of Web Search. In *2019 Conference on Human Information Interaction and Retrieval (CHIIR '19), March 10–14, 2019, Glasgow, United Kingdom.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3295750.3298922

1 INTRODUCTION

During an information seeking episode, people often engage in a series of information behaviors with various goals and intentions which emerge from characteristics of their tasks and problematic situations. In the context of interactive information retrieval (IIR), a user's information seeking tactics are generally motivated and heavily influenced by their *information seeking intentions* [19]. Previous research has shown that people intend to achieve different goals and thereby adopt different search strategies at different stages during the course of information search [22, 27]. To design an adaptive IR system and to support users' interactions with information, IR researchers need to embrace a comprehensive understanding of how search task and search behaviors are connected to information seeking intentions at different stages of a search session.

According to the cognitive viewpoint of IR research, a search task can be conceptualized as a sequence of information seeking intentions which both motivate and are influenced by Web search behaviors. Previous IR literature on Web search has discussed the effects of task features on users' search behavior and experience (e.g., [1, 10, 16, 17]). While the findings from this line of research imply the cognitive changes behind the tasks and behavioral variations in search, the relationships between task, information seeking intentions, and user behavior still remain unexplored. To address this research gap and to model the interactions between different levels (i.e., task level, intention level, and behavior level) of Web search, IR researchers need to directly collect data on the various goals that users attempt to achieve in local steps (i.e., query segment) of task-based information seeking and search session, and to analyze: 1) how users' information seeking intentions vary across different query segments and task types; 2) how users' information seeking intentions are associated with their Web search behaviors.

To address this issue, the authors conducted a controlled lab study in which participants were assigned two different work tasks, and to accomplish a specific, different search task for each. After each search session, a recording of the session was played back, and participants were asked to explicitly annotate their information seeking intentions for *each query segment*. In this study, query segment is defined as *a search session segment which starts from one*

issued query and ends at the next query. Based on the data collected from 693 query segments generated by 40 participants in 80 search sessions, the authors analyzed the effects of search task features on the frequency of occurrence of information seeking intentions in different query segments/stages of search, and also examined the relationships between Web search behavior and information seeking intentions in current and subsequent query segments respectively.

This study has multiple contributions: 1) Overall, this study reveals the connections between task, information seeking intentions, and user behavior in Web search and enriches the multi-level understanding of users' interaction with IR systems. 2) With respect to the effects of task features, this study offers direct evidence on the variations in users' information seeking intentions in tasks of different types and demonstrates that some of the task effects on intentions vary significantly across different query segments within a search session. 3) Regarding the relationships between intention and user behavior, this study indicates that the presences of a variety of information seeking intentions in both current and subsequent query segments are closely connected with and detectable by different subsets of behavioral measurements. 4) From the IIR system design and evaluation perspective, this work has implications for system designers on how to better support different information seeking intentions and facilitate users' interactions with information at different stages of search and in different task scenarios. Also, this study encourages future researchers to develop new evaluation metrics and frameworks tailored to different types of intentions.

2 BACKGROUND

2.1 Information Seeking Intentions

Search task is a complex, multidimensional concept. To decompose search task, prior studies mainly focused on a set of widely-discussed search task features (e.g., task complexity, task difficulty, task type) and their connections to the variations in Web search behavior (e.g., [1, 12]). Regarding the cognitive space and intention aspect of users, however, information seeking intentions in query segments as cognitive-level components of search task have been scarcely studied. As a result, although users often navigate to useful information with small, local steps under the influence of search tasks [25], it is still unclear how their information seeking intentions in local steps (i.e., query segments) are connected to low-level search behavior and high-level, global context (i.e., search task).

Some early theoretical research on classifying search sessions could be thought of as applicable to information seeking intentions. For instance, Broder [3] argued that Web searches in terms of information seeking and search intention could be classified into three categories: navigational, transactional, and informational. Aiming to build a more detailed scheme of intentions, Kellar, Watters, and Shepherd [11] proposed a new typology of user intentions generated in search, which includes fact finding, information gathering, browsing, and transactions. These classifications, as is some later work discussed below, were built upon analysis of queries that initiated search sessions and did not take into account the search actions within query segments. Also, these early theoretical classifications only identified broad intention categories and ignored the nuance between different specific intentions within each category.

Besides the theoretical speculation on the classification of intentions, according to Rha et al. [22], Xie [27] is the only example of an empirically-based classification of intentions which motivate people to engage in different interactions with search engines. Nevertheless, other similar research on users' goals, knowledge gap, and search intent can also be of help to understand users' intention in information seeking and search episodes. For example, Rose and Levinson [23] analyzed a set of queries randomly selected from AltaVista query logs and proposed a hierarchical typology of users' search goals. Similarly, drawing on the ideas of the Sense-Making approach [7], Savolainen and Kari [24] revealed the discontinuous and dynamic nature of Web searching episodes and developed a conceptual framework of knowledge gaps faced by searchers as well as the corresponding gap-bridging strategies. Jansen and Booth [9] developed a three-level hierarchy of user intent to automatically classify Web search queries based on the information seeking intentions behind these queries. They found that users' query intent (i.e., informational, navigational, transactional) varies by different search topics. In recent research on information seeking intention, Mitsui, Shah, and Belkin [21] developed a set of information seeking intentions based on Xie [27]'s initial typology of interactive intentions and empirically investigated the distributions of different intentions in search tasks of different types. Rha et al. [22] studied how different types and states (i.e., satisfied or unsatisfied) of information seeking intentions lead to different query reformulation strategies. Despite the increasing attention to the role of user intentions, the connections between task, information seeking intentions, and user behavior still has not been systematically studied.

2.2 Task and User Behavior in Web Search

In much of previous research which uses the concept of task, a task was conceptualized either as an abstract construction or as a concrete sequence of actions [4]. Following this line of thought, a large body of IIR research has explored the relationships between task level and behavior level of Web search and investigated the behavioral effects of various task features which include both objective task facets (e.g., task goal, task product) as well as user-task combined features (e.g., perceived task difficulty, topic familiarity). For instance, Jiang, He, and Allan [10] examined the associations between user behaviors (i.e., searching, browsing and clicking) and objective task features (i.e., task product, task goal, task type) and discussed to what extent these behavioral features can help disambiguate search tasks of different types. Capra et al. [5] found that manipulating task priori determinability via modifying task items and dimensions can significantly affect users' task perceptions and choices of search strategies. Similarly, Liu et al. [17] indicates that some whole-session behavioral metrics and eye movement features can serve as useful indicators of search task traits.

With respect to task-user combined features (i.e., users' knowledge and perception concerning tasks at hand), Wildemuth [26] argues that in task-based information search, users' search tactics are influenced by their domain knowledge related to the task topics. Liu, Gwizdka, Liu, and Belkin [18] demonstrates that both whole-session level and within-session level behaviors are affected by task difficulty, and that the dynamic relationships between search behavior and task perception are subject to the influence of task type

(i.e., single fact-finding, multiple fact-finding, and multiple-piece information gathering). Similarly, Aula, Khan, and Guan [1] also investigated search behavioral variations under task of different levels of difficulty. By conducting both a lab study and a large-scale online study, they found that when performing difficult search tasks, users tend to issue more diverse queries (have a more unsystematic query refinement process), use advanced operators more frequently, and spend longer time on search engine result pages (SERPs) during their search processes. With respect to the facet of time pressure, Liu and Wei [15] found that users tend to become more selective in examining search results when searching under time pressure.

In addition to the explorations on task-behavior interactions, IR researchers also sought to develop user models based on large-scale behavior log data (e.g., clickstream data, query logs). For example, Borisov et al. [2] constructed a click sequence model (CSM) to predict the order in which a user will interact with the search results on SERPs. Hassan, Jones, and Klinkner [8] conducted behavior analysis based on 1000 user sessions extracted from Yahoo! search engine logs and found that user behavior can serve as a better predictor than document relevance on predicting search success.

Although the previous research discussed above enriched our understanding of Web search interaction and its associations with the overarching tasks, it did not offer us a more holistic understanding of the relationships between task as global context, information seeking intentions in local steps, and user behavior. Especially, the role of intentions in bridging global search context and local search tactics still remains ambiguous.

3 RESEARCH QUESTIONS

To address the research gaps discussed above and to develop a multi-level understanding of Web search, this work reports on a controlled-lab-based study which investigated information seeking intentions and search activities in query segments generated by forty participants in tasks of different types. The authors sought to answer the following research questions (RQs):

- **RQ1a**: What are the effects of task features on the frequency of occurrence of different information seeking intentions in query segments?
- **RQ1b**: To what extent do the frequency of occurrence of different information seeking intentions vary across different stages (query segments) of a search session?
- **RQ1c**: To what extent do the effects of task features vary across different stages of a search session?
- **RQ2a**: How is a user's search behavior associated with her information seeking intentions in the same query segment?
- **RQ2b**: How is a user's search behavior in the current query segment associated with her information seeking intentions in the next/subsequent query segment?

Among the RQs above, RQ1a-c speak to the effects of task features on information seeking intentions in query segments, and RQ2a-b were proposed to examine the relationships between Web search behavior and information seeking intentions at different stages. The lab study designed for answering the research questions is described in detail in the following methodology section.

4 METHODOLOGY

4.1 Information Seeking Intentions

To classify users' information seeking intentions in Web search query segments, this research used the typology of search intentions which was developed and elaborated by Rha et al. [22] based on a subset of Xie [27]'s classification of interaction intentions. The authors gave a description of this typology to the participants before their search sessions were replayed for intention annotation. Then, participants were asked to identify their information seeking intention(s) for *each query segment* based on the typology. Participants could identify multiple intentions in the cases where they tried to accomplish multiple things within a single query segment. Therefore, the intentions identified in our study are not mutually exclusive and should be considered as different variables. Details regarding the intention annotation process are provided in "Procedures" section below. The twenty information seeking intentions are listed in Table 1.

4.2 Search Task and Participants

To control the task context, the authors designed four work tasks within the domain of journalism: Copy editing (CPE), Story pitch (STP), Relationships (REL), and Interview preparation (INT). Forty undergraduate students majoring in journalism were recruited as participants from a research university. To ensure that the participants have a certain level of familiarity with journalism tasks, only upper-division undergraduates who have completed either one journalism writing or reporting class were selected.

The four different search tasks types were defined using four task facets extracted from the task classification scheme proposed by Li and Belkin [14], as modified by Cole et al. [6]: Product, Goal, Level, and Named. With respect to task Product, intellectual task refers to a task which produces new ideas or findings, whereas factual information task refers to a task locating facts, data, or other similar information objects in IR systems. Regarding task Goal, task with specific goal refers to a task with a goal that is explicit and measurable. By contrast, task with amorphous goal is defined as a task with a goal that has no explicitly defined outcome which cannot be measured in a quantitative sense. In terms of Level, a task can be classified into two categories: 1) document-level task: a task for which a document as a whole is judged; 2) segment-level task: a task for which a part or parts of document are judged. The Named facet refers to whether what is to be found is explicitly named; its values are true or false, named or not named. The values of each value for each of the task types are shown in Table 2.

Each of the four task types has two specific versions, which differ from one another in topic (i.e., coelacanths; methane clathrates and global warming). These two topics were selected to control the variable of participant familiarity with the task topic, as our participant population was thought likely to not be familiar with either. Table 2 shows the four specific tasks for the topic of coelacanths; "methane clathrates and global warming" was substituted for "coelacanth" to generate four more specific tasks. Each of the participants was asked to do two tasks of different topics and types, in Latin Square design, pairing CPE with INT, and STP with REL, to balance tasks by topic, and by the values of Product, Level and Named facets.

Table 1: Information seeking intentions and the associated acronyms used in Table 6, 8 and 9

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

Table 2: Tasks assigned in lab study

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

4.3 Procedure

The lab study started with a demographic questionnaire and a tutorial video on the additional browser interface features our plug-in provided for participants in information search sessions. Participants were free to search anywhere on the web, with the only restriction being to conduct their interactions within the supplied browser, which logged their search behaviors, and provided the ability to save and unsave web pages. The participants then went through the task description (see Table 2) and answered a short questionnaire on task familiarity/experience, topic familiarity, and anticipated task difficulty. They then had up to 20 minutes to complete the first assigned search task, by searching anywhere on the Web, with any search facilities that they wished, but could choose to finish early if they completed it to the best of their ability. Participants' search activities (e.g., query, timestamp, URLs, page type) were recorded with a Firefox browser plugin and Morae (https://www.techsmith.de/morae.html).

Afterwards, participants were asked to finish a post-search questionnaire on search task difficulty and self-reported search experience. At this point, they were asked to read a guidance of the intention annotation task and to watch a video explicating how to annotate search intention in each query segment. The authors then replayed the entire search, query segment by query segment, asking for intention annotation for each segment, in sequence. In our study the intention annotation task had no time limit. For the intention annotation task, participants were asked to select which intentions applied to each query segment in the search session. Within each query segment, participants could choose any number of intentions from the list. Participants could choose "other" if their intention did not match the 20 intentions provided. Participants repeated this selection process for each query segment during the course of intention annotation. The same procedure was then followed for the second search task, and the study session ended with an exit interview with questions related to participants' experience of performing the two search tasks. The entire process took about two hours for each participant.

4.4 Measurements and Data Analysis

With respect to task-intention connection, to answer RQ1a, the authors defined the binary status (i.e., present or absent) of information seeking intentions as dependent variables and employed task type (i.e., copy editing, story pitch, relationship, interview preparation) and task-user combined features (i.e., task familiarity, topic familiarity, perceived task difficulty, time condition/perceived time pressure) as independent variables and built logistic regression models for different information seeking intentions respectively. Data concerning the independent variables was collected via pre- and post-search questionnaires. To clarify the actual meaning of these variables, this section lists the specific questions corresponding to the independent variables in Table 3.

In addition to the task variables, to answer RQ1b, the logistic regression models also include *query percent* (QE) (query percent = n/N, where n is the sequence number of the query segment (nth query segment)) within the corresponding search session and N represents the total number of query segments within the search session) as an independent variable to describe the sequential aspect of query segments. Query percent was measured at the end of each query segment and indicates the stage that the user was in within a given search session. For example, in a search session consisting of four queries, the query percent value of the first and second

Table 3: Task-user combined features

Variables	Questions and Scales
Task familiarity	How much experience do you have with this kind of assignment? 1 (not at all)-7 (extremely) - pre-search
Topic familiarity	How familiar are you with the topic of this assignment? 1 (not at all) - 7 (extremely) - pre-search
Task difficulty	How difficult was it to find the information you need for this assignment? 1 (not at all) - 7 (extremely) - post-search
Time condition	Did you have enough time to complete the assignment successfully? 1 (far too little), 2 (too little), 3 (barely enough), 4 (enough), 5 (more than enough) - post-search

query segments are 0.25 and 0.5 respectively. However, in a search session consisting of only two queries, the query percent of the first and second query segments are 0.5 and 1 respectively. Using query percent instead of the sequence number of a query can make two distinct query segments from different sessions more comparable.

To answer RQ1c, the authors included the interaction between query percent and the other independent variables in the regression models, aiming to examine to what extent the effects of those variables on users' intentions change across different query segments.

To answer RQ2a and RQ2b, this study employed seven Web search behavioral measurements calculated based on the search log data, addressing different aspects of Web search (see Table 4). This study used Mann-Whitney U test to examine intention-behavior connection as none of the search behavior data was normally distributed. The statistical test for each intention-behavior pair was conducted between intention selected group and the entire sample. Specifically, in data analysis, for each type of intention, the authors first statistically tested the difference between the mean of value ranks of each search behavior data of users who had the intention in *current query segment* and that of the same search behavior of the entire sample (RQ2a). Similarly, to respond RQ2b, the authors also tested the difference between the mean of value ranks of each search behavior data of users who had the intention in the *next/subsequent query segment* and the mean of value ranks of the corresponding search behavior of the entire sample in current query segment, aiming to test the connection between current search behaviors and subsequent intentions.

Table 4: Behavioral measurements

Type	Behavioral Measurements
Query behavior	query length (number of terms)
Browsing behavior	number of clicks, number of sources (i.e. unique Internet domains) visited, number of pages (i.e., including repeated pages) viewed
Dwell time (second)	mean dwell time on each SERP, mean dwell time on each content page
Useful judgment	number of bookmarks

5 RESULTS

5.1 Descriptive Statistics

To answer the research questions proposed in previous sections, the authors conducted statistical analysis based on the intention and search behavior data collected from 693 query segments generated by 40 participants in tasks of four different types. For each intention, Figure 1 reports the proportions of query segments in which the intention was selected by users (which means the users had this intention in the corresponding query segments). The result shows that except for the intention of finding specific information, for almost every information seeking intention, the proportions of query segments where the intention was selected is smaller than that of the opposite cases (only three intentions appeared in more than 30% of query segments). This result indicates that there is no ubiquitous information seeking intention in Web search. Instead, the presences of most information seeking intentions may depend on a series of contextual factors and restrictions (e.g., stage of search, task type, users' perception of task). To test the effects of task features (including both objective task type and subjective task-user combined features) on information seeking intentions and to answer RQ1a-c, the authors conducted logistic regression analyses and the results are reported in the following section.

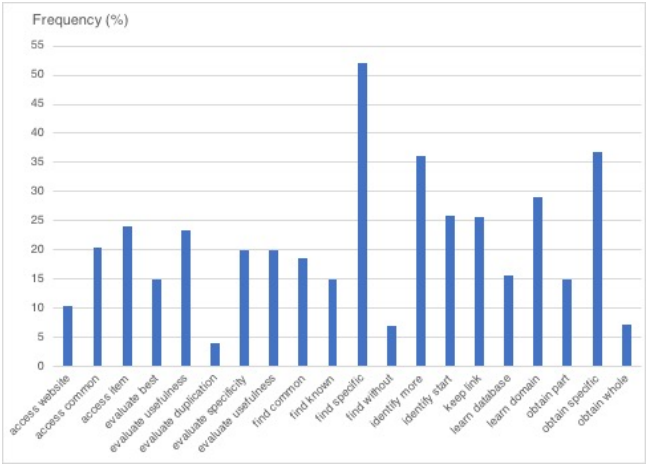

Figure 1: Intention selected in 693 query segments (%)

Table 5: Descriptive statistics of independent variables

Variable	Mean	S.D.	Min.	Max.
Session length	8.66	5.83	2	29
Task familiarity (TE)	3.04	1.67	1	7
Topic familiarity (TF)	1.88	1.34	1	6
Task difficulty (TD)	2.66	1.58	1	7
Time condition (TC)	4.18	.93	2	5

With respect to other independent variables included in logistic regressions, Table 5 reports the descriptive statistics of search session length (measured by number of queries) and task-user combined features. Search sessions which include 3-7 query segments appeared more frequently than other search sessions of other

lengths. The relatively large variance in search session length (Mean = 8.66, S.D.=5.83) across different sessions demonstrate that query segments with the same sequence number may represent largely different stages of their respective sessions. Thus, compared to the sequence number of query segments, *query percent* may be a better measurement to describe which search stage a user was in.

For task familiarity (TE), topic familiarity (TF), and task difficulty (TD) measures, the mean values are more closer to the left ends of the scales. Particularly, for topic familiarity measure, the mean value is quite low and in more than 60 of 80 sessions participants identified their level of topic familiarity as 1, indicating that most of the participants were not familiar with the topics at all. This result demonstrates that the topic selection was successful and served the original purpose. The mean value of time condition (TC) score is relatively high, indicating that most of the participants had enough time to complete their tasks. In this time-sufficient setting, participants were more likely to behave in their natural manner.

5.2 From Task to Intention: RQ1a-c

To answer RQ1a-c, this section presents the results of logistic regression analysis for every information seeking intention. Tables 6 report the coefficients of different independent variables (IVs) (task type, task-user combined features, query percent, and interaction terms) in each logistic regression model. This model included three binary variables (STP, REL, INT) to represent the categorical variable task type. The copy editing task (CPE) was not explicitly included and thus was used as a baseline against which the other task were compared because CPE as a factual specific task has a lower level of cognitive complexity compared to other task types [12]. Thus, CPE as a baseline can better help explain the underlying connection between the cognitive demands of different tasks and the frequency of occurrence of different intentions.

Among the reported coefficients, the coefficients of task features can help explain the effects of these variables on the frequency of occurrence of different information seeking intentions (RQ1a). The coefficients of query percent and interaction terms can clarify to what extent the frequency of occurrence of intentions and the main effects of task type and task-user combined features varied across different query segments in search sessions (RQ1b and RQ1c).

According to the results reported in Table 6, the coefficient of query percent is negative and statistically significant in the “AW" column, indicating that as search session proceeded, participants became less likely to directly go to a website or homepage. This result suggests that participants tended to access a known website or homepage at early stage of search session as it could be an easy and useful way to start searching and dealing with the anomalous state of knowledge at the beginning stage. The results in the “AC" column demonstrate that participants with higher level of topic familiarity were more likely to access items with common characteristics. This may be because high level of topic familiarity decreased the difficulty and cognitive load of finding items with common features, and thereby encouraged participants to seek similar items for resolving the tasks. The results reported in the “AS" column indicate that participants accessed specific known item less frequently in story pitch tasks than in copy editing tasks, indicating that it might be difficult (or not necessary) for participants to use a known item(s) in a factual task without a clearly defined goal.

With respect to the intentions under *Evaluate* category, because of the nature of story pitch task (find webpages which contain the most interesting facts about the topic), participants who conducted this task tended to spend more query segments on picking best item(s) from all the useful ones. The intention of evaluating correctness of an item was affected by both the main effects of task features and an interaction term (task familiarity x query percent). Specifically, when participants were conducting story pitch and interview preparation tasks, they were less likely to evaluate the correctness of retrieved items in query segments as these tasks were open-ended and did not require participants to confirm or disconfirm any statement. Interestingly, the results indicate that higher task familiarity led participants to use significantly more query segments in evaluating correctness. However, as the search sessions proceeded, the effect of task familiarity decreased over time (the coefficient of the interaction term is a negative value). Despite the positive effect on the frequency of occurrence of the intention of evaluating correctness, higher task familiarity was associated with significantly lower frequency of evaluating duplication of the retrieved items. The intention of evaluating specificity appeared less frequently in the story pitch task compared to the “baseline" copy editing task as participants needed to find plenty of relevant, specific items to finish copy editing tasks (confirm or disconfirm the given statements). The results presented in “EU" column indicate that participants’ intention of evaluating usefulness was affected by both task type and task-user combined features (topic familiarity and task difficulty). Whereas the positive effect of relationship task increased as the search session proceeded, the effects of both topic familiarity and task difficulty gradually decreased over time.

With respect to the intentions under the *Find* category, the frequencies of occurrence of different intentions were associated with different sets of factors. When a participant was more familiar with the task at hand, he or she was less likely to spend time on finding items with something in common. This may be because when performing familiar tasks, participants preferred to search for more diversified results and do more explorations, instead of relying on similar websites and documents. In contrast, a high level of topic familiarity encouraged participants to seek similar items. However, this positive effect of topic familiarity decreased as the search session proceeded. When under less time pressure (having enough time to complete the tasks), participants were more likely to spend time on seeking known items. Similar to the aforementioned effect of topic familiarity, this positive effect of time condition also decreased over time (the coefficient of the interaction term TC*QP is a negative value). In terms of the intention of finding specific items, compared to the copy editing task which requires more specific, predefined information, interview preparation as an open-ended task were associated with significantly lower frequency of seeking a predetermined piece of information. In addition, a higher level of task familiarity also was associated with a decrease in the occurrence of the intention of finding specific items. The intention of finding items without predefined criteria frequently happened in tasks with a high level of difficulty as difficult tasks involved more exploratory search activities (without predetermined guidance). In addition, participants who had sufficient time to work on their tasks

Table 6: The effects of task type and task-user combined features on intentions: Logistic regressions (*p<.05, **p<01. light grey: positive effect. dark grey: negative effect. white: no significant effect). IV (independent variable) definitions are in the Appendix Table 7; Table 1 gives the meaning of the intention acronyms (column labels).

IVs	AW	AC	AS	EB	EC	ED	ES	EU	FC	FK	FS	FW	IM	IS	KR	LD	LK	OP	OS	OW
STP	-1.1	-.97	-1.6*	.83*	-.14*	.87	-1.3*	-.33	.03	.10	-.62	.46	.28	-.05	-.68	-.52	.57	-.78	-1.5*	.02
REL	-.28	.84	.22	-.01	-.04	.28	.02	1.1*	1.1	.26	-.14	-.05	1.2*	-.05	-.13	-1.2	1.1*	-.13	-.37	.14
INT	-.84	.12	-.01	-.16	-2.6**	-.58	-.26	-.96	.73	-.36	-.11*	-.50	1.1*	.31	-.69	-.67	.55	-.09	-.76	-.65
QP	-1.9*	.81	1.3	.30	1.3	2.7	1.4	.05	.76	3.5	-3.9	3.5	-.31	-.98*	-.94	-8.6**	2.1	-3.9	3.4	-5.6
TE	.09	-.26	.13	-.04	.45**	-.48**	-.01	.17	-.42**	.03	-.26**	.06	-.47	-.34**	-.17	-.18	-.17	-.22	.07	-.27*
TF	.08	.30*	.09	.02	-.23	.18	.04	-.26*	.53**	.26	.17	-.12	.12	-.29*	.22	.14	.41**	.48**	-.13	-.14
TD	.17	.23	.25	-.14	.35	.29	.22	.33*	.23	.10	.08	.53*	-.01	.09	.36*	-.33	.25	.20	.43**	-.32
TC	.34	.40	.46	.26	.18	.57	.37	.42	.20	.51*	.18	.92*	.15	.21	.52*	-.59	.76**	.49	.69**	.06
STP*QP	.72	.71	1.0	-.85	-.27	-1.7	.83	1.5	-.27	-1.7	-.12	3.9	.14	-1.7	.99	.25	-.58	1.0	1.2	1.5
REL*QP	.79	.57	-.64	.16	-1.1	-.78	.33	2.5*	-.17	-.36	-.03	3.1	-1.5	-.66	-.02	2.2	-.52	.28	.20	.65
INT*QP	.08	.41	-.46	.79	.93	.10	-.07	3.2	-.52	-.28	1.3	4.5	-1.0	-2.1	.90	-.44	-.82	-.42	.15	1.1
TE*QP	-.01	.31	.13	.11	-.61*	.79	-.13	-.45	.34	.48	.19	-.18	.28	-.57	-.17	.17	.09	.30	.05	.67
TF*QP	.25	-.19	.15	.28	.57*	-.36	.22	.64*	-.73*	-.24	-.09	.05	-.13	.61	.06	.04	-.46	-.11	.26	.24
TD*QP	.01	-.74	-.52	-.41	.35	-.79	-.61	-.78*	-.65	-.33	.14	-1.2*	-.13	-.01	-.55	.29	-.09	.25	-.77*	.42
TC*QP	.19	.01	-.04	-.23	.15	-.51	.01	.17	.44	-1.0*	.64	-1.1	.031	.29	.24	1.6	-.34	.36	-.64	.01

were more likely to do exploratory searches and seek for useful webpages without predetermined features. Again, this main effect gradually decreased as the search session proceeded.

In terms of the intentions under *Identify search information* category, participants were more likely to identify something more to search and to explore a topic more broadly in the two open-ended tasks (relationship task and interview preparation task) than in the copy editing task. With respect to the intention of identifying something to get started (e.g., find good query terms), it appeared more frequently in the earlier stages than in the later stages (the coefficient of query percent is a negative value). Besides, high levels of task familiarity and topic familiarity significantly decreased the frequency of occurrence of this intention, indicating that when a participant was familiar with the task or topic, he or she was more likely to quickly find useful documents and thus did not need to frequently try and restart new search paths.

Based on results in the "KR" column, when a participant was performing a difficult task or having sufficient time in search, he or she tended to save more useful items to look at later. Regarding the intentions under the *Learn* category, participants tended to learn more about the types of available resources at particular websites in the early stages of search than in later stages. Participants usually spent more query segments on learning domain knowledge in the relationship task than in the copy editing task. Besides, having high level of topic familiarity or sufficient time in search usually led to higher frequency of seeking domain knowledge.

With respect to the intentions under the *Obtain* category, when a participant was familiar with the topic of the assigned task, he or she was more likely to seek for parts of relevant items to highlight or copy. A possible reason is that the familiarity of topic could help participants to easily locate useful pieces of information within retrieved items. Compared to the baseline situation (copy editing task), participants who were doing the story pitch task were less likely to seek for specific information to highlight or copy. Participants tended to seek specific known items when they experienced higher task difficulty or had relatively sufficient time in search. However, the positive effect of task difficulty tended to decay over time in search sessions. Participants who were doing familiar tasks were less likely to seek for a whole item to highlight or copy. This may be because obtaining a whole item is an inefficient way of seeking useful information and thus participants who were familiar with their tasks preferred to avoid this inefficient approach.

5.3 From Intention to Behavior: RQ2a-b

To answer RQ2a-b, this section presents the results of Mann-Whitney U test for each intention-behavior pair (see Table 8 and 9 in the Appendix). Specifically, Table 8 illustrates the connection between current search behaviors and current information seeking intentions (RQ2a), and Table 9 shows the relationship between current search behaviors and information seeking intentions in the next/subsequent query segment (RQ2b). Note that we did not employ Kruskal-Wallis test or the associated post hoc pairwise tests (e.g., Bonferroni test) here as the information seeking intentions were not mutually exclusive (i.e. users can select multiple intentions for the same query segment if applicable).

Overall, according to the results in Table 8, all intentions identified were significantly associated with at least one aspect of Web search in current query segment, indicating that information seeking intentions were connected to and reflected by Web search behavior in various ways. For intentions under *Access* and *Evaluate* categories, they were mostly correlated with click activity, dwell time on SERP, and usefulness judgment behavior. Specifically, when participants had intention(s) related to accessing item(s) or evaluating obtained information, they tended to be more active in browsing, page viewing, and judging usefulness of gathered information. In particular, when the identified intention(s) was concerned with comparisons between different results (e.g., access items with common features), participants spent more time on evaluating the results on SERPs visited. As a result, in "AC" and "EB" columns, SERP dwell time was significantly higher than the average level of the sample.

In contrast, intention under *Find* category did not show any unified pattern in intention-behavior connection. In other words, different specific intentions in this group were associated with different intentions. For example, the intention of finding information with common features was positively associated with SERP dwell time as this intention involves evaluating and comparing results on

SERPs. However, the intention of finding specific information was more closely correlated with content page viewing and usefulness judgment behaviors. With respect to the *Identify* category, participants tended to leave SERPs earlier when they merely wanted to identify something more to search, rather than dig deep into the results. When participants sought to identify a starting point for Web search, they were more active in content page reading and clicking, but less patient with SERP viewing and visiting more pages.

Similar to the intentions in *Access* and *Evaluate* groups, most of the intentions under *Keep, Learn* and *Obtain* categories were also closely related to clicking, content page viewing, and bookmarking behaviors, indicating that intentions under these categories could lead to similar search tactics. Among these intentions, when participants sought to learn domain knowledge, they tended to spend less time on browsing SERPs but much longer time on carefully reading the clicked content pages. The intentions of learning database information and obtaining a whole item(s) were only associated with one of the search behaviors, suggesting that their connections to Web search might be loose. Another possible reason is the small scale of sample size. Neither of these two intentions occurred frequently in the collected query segments (frequencies were both lower than 15%). Consequently, the limited sample size may not be able to fully picture the connections between intentions and behaviors.

With respect to the RQ2b, compared to the rich connections between search behavior and current intentions, fewer significant associations were found between search behavior and subsequent intentions (see Table 9). Among all types of intentions, seven intentions (i.e., access specific item(s), evaluate correctness, evaluate duplication, find specific information, learn database information, obtain specific information, obtain a whole item to highlight or copy) had no significant connection with search behavior. This result indicates that some of the intentions were more closely restricted to the local steps or micro-situation (i.e., query segment) and were loosely connected (if not totally disconnected) with previous actions. For other intentions, most of them were positively associated with click activity and usefulness judgments (i.e. bookmarking) behaviors. For example, when participants were active in clicking results and seeking useful webpages, in the subsequent query segment they were more likely to engage in a series of intentions concerning with accessing item, evaluating information, keeping records of useful links, and learning domain knowledge. In addition, when users issued a short query and had a relatively broad search scope in the current query segment, they tended to continue their exploratory search intentions in the subsequent query segments, such as evaluating specificity of obtained items, identifying more relevant information to continue search, and seeking information without predefined criteria. The intentions of finding information with common features and finding known item(s) were only associated with clicking, suggesting that active clicking in current query segment as sunk cost could encourage participants to seeking more gains (i.e., finding more information) in the next round.

6 DISCUSSION AND CONCLUSION

To explore the connections between global contexts and local steps in IIR and to develop a multi-level understanding of Web search, this study examined the effects of task features on information seeking intentions at different stages of search (RQ1a-c) and investigated the connections between search behavior and intentions in both current and subsequent query segments (RQ2a-b). With respect to the research questions, the authors have following answers.

In terms of the effects of task features on user intentions, for RQ1a, our results demonstrate that search task types and user characteristics significantly affected the frequency of occurrence of 18 out of 20 information seeking intentions (excluding only access website and learn database). It is worth noting that task type and task-user combined features (i.e. topic familiarity, perception of task) generated different effects on different information seeking intentions. These results supplement the findings on search behaviors from previous IIR research (e.g., [10, 17]) by confirming the effects of task features on information search at intention level.

To answer the RQ1b, the authors examined the main effect of query percent on the occurrence of different intentions. As search sessions proceeded, participants became less likely to access a particular website or homepage, identify something to start searching, or learn about the resources in databases. For other intentions, the main effect of query percent was not significant, meaning that the local context of search did not significantly affect these intentions. These results enrich our understanding of the cognitive variations in Web search and empirically confirm the classical arguments that users have different intentions at different stages of information search process [13]: for instance, users tend to do more exploratory searches at the beginning and then decrease these open-ended attempts once they formulate their search focuses.

For the RQ1c, results indicate that as the search sessions proceeded, many of the main significant effects of task features gradually faded away. This finding suggests that the main effects of task features are usually stronger at the early stages (the initial query segments) of search. Thus, early stages of a search session can be a good time for task prediction. This result has implications for task modeling and helped explain Mitsui, Liu, and Shah [20]'s findings: first query measures can be at least as good as, and sometimes better than, whole session measures for task type predictions.

With respect to RQ2, aside from clarifying the connections between intentions and search behavior, the results summarized in Table 8 and 9 also have implications for prediction analysis. Each column represents a vector of significant correlations associated with the corresponding intention. For instance, the behavioral vector for the current intention of identifying something to start is $[clicks = above, sources = below, pages = below, SERP = below, content = above]$. To accurately identify users' intentions in the current query segment and more importantly to predict their intentions in the subsequent query segment, the ideal situation is that we could find a unique vector of behavioral measures for each information seeking intention. In this study, the authors identified unique behavioral vectors for 9 current intentions and 7 subsequent intentions. Future studies need to include more behavioral measures and identify unique behavioral vectors for other intentions.

With respect to IIR system design and evaluation, by highlighting the role of information seeking intentions in connecting task features and search behaviors, this study calls for future research efforts on designing new system affordances and identifying more fine-grained system evaluation metrics tailored to different types of intentions in different search stages and task contexts.

7 ACKNOWLEDGEMENTS

This work is supported by the National Science Foundation (NSF) grant IIS-1423239.

REFERENCES

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

[2] Alexey Borisov, Martijn Wardenaar, Ilya Markov, and Maarten de Rijke. 2018. A Click Sequence Model for Web Search. *arXiv preprint arXiv:1805.03411* (2018).

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

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

[5] Robert Capra, Jaime Arguello, Heather O'Brien, Yuan Li, and Bogeum Choi. 2018. The Effects of Manipulating Task Determinability on Search Behaviors and Outcomes. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. ACM, 445–454.

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

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

[8] Ahmed Hassan, Rosie Jones, and Kristina Lisa Klinkner. 2010. Beyond DCG: user behavior as a predictor of a successful search. In *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 221–230.

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

[10] Jiepu Jiang, Daqing He, and James Allan. 2014. 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*. ACM, 607–616.

[11] Melanie Kellar, Carolyn Watters, and Michael Shepherd. 2007. A field study characterizing Web-based information-seeking tasks. *Journal of the American Society for Information Science and Technology* 58, 7 (2007), 999–1018.

[12] Diane Kelly, Jaime Arguello, Ashlee Edwards, and Wan-Ching Wu. 2015. 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*. ACM, 101–110.

[13] Carol Collier Kuhlthau. 1999. The role of experience in the information search process of an early career information worker: Perceptions of uncertainty, complexity, construction, and sources. *Journal of the American Society for information Science* 50, 5 (1999), 399–412.

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

[15] Chang Liu and Yiming Wei. 2016. The impacts of time constraint on users' search strategy during search process. In *Proceedings of the 79th ASIS&T Annual Meeting: Creating Knowledge, Enhancing Lives through Information & Technology*. American Society for Information Science, 51.

[16] Jingjing Liu and Nicholas J Belkin. 2010. 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*. ACM, 26–33.

[17] Jingjing Liu, Michael J Cole, Chang Liu, Ralf Bierig, Jacek Gwizdka, Nicholas J Belkin, Jun Zhang, and Xiangmin Zhang. 2010. Search behaviors in different task types. In *Proceedings of the 10th annual joint conference on Digital libraries*. ACM, 69–78.

[18] Jingjing Liu, Jacek Gwizdka, Chang Liu, and Nicholas J Belkin. 2010. 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*. American Society for Information Science, 16.

[19] Matthew Mitsui, Jiqun Liu, Nicholas J Belkin, and Chirag Shah. 2017. Predicting Information Seeking Intentions from Search Behaviors. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 1121–1124.

[20] Matthew Mitsui, Jiqun Liu, and Chirag Shah. 2018. 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*. ACM, 1141–1144.

[21] Matthew Mitsui, Chirag Shah, and Nicholas J Belkin. 2016. Extracting Information Seeking Intentions for Web Search Sessions. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 841–844.

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

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

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

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

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

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

8 APPENDIX

Table 7: Acronyms of the independent variables in the logistic regressions and the definitions

Acronym	Definition
IV	Independent variable
STP	Story pitch task (binary, Y=1, N=0)
REL	Relationship task (binary, Y=1, N=0)
INT	Interview preparation task (binary, Y=1, N=0)
QP	Query percent
TE	Task familiarity/experience
TF	Topic familiarity
TD	Task difficulty
TC	Time condition
STP*QP	The interaction term between STP and QP
REL*QP	The interaction term between REL and QP
INT*QP	The interaction term between INT and QP
TE*QP	The interaction term between TE and QP
TF*QP	The interaction term between TF and QP
TD*QP	The interaction term between TD and QP
TC*QP	The interaction term between TC and QP

Table 8: Median of behavioral measure when the intention was selected in *current* query segment. (Mann-Whitney test. light grey: above the mean of total ranks. dark gray: below the mean of total ranks. white: no significant difference.)

Behavior	AW	AC	AS	EB	EC	ED	ES	EU	FC	FK	FS	FW	IM	IS	KR	LD	LK	OP	OS	OW
query	4	4	4	4	4	3	4	3	4	4	4	3	4	4	4	4	4	4	4	4
click	4	3	4	5	3	6	4	4	3	4	3	3	3	3	4	3	3	4	3	3.5
source	4	4	4	4	4	5	4	4	4	4	4	4	4	4	5	4	4	4	4	4
page	5	5	5	5	5	8	5	5	5	5	5	5	5	4	6	5	5	5	5	5
SERP	8.2	9.5	7.1	7.9	9.1	4.8	8.4	8.1	7.7	6.2	8.8	7.3	6.7	7.4	8.8	7.5	7.3	12.4	9.6	6.8
content	11.4	11.6	11.8	11.9	11.8	17.7	13.2	13.9	9.5	10.7	11.5	14.0	9.9	11.7	13.9	7.7	12.8	12.4	13.4	14.6
bookmark	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	0	1	1	1	1

Table 9: Median of behavioral measure when the intention was selected in *next* query segment. (Mann-Whitney test. light grey: above the mean of total ranks. dark gray: below the mean of total ranks. white: no significant difference.)

Behavior	AW	AC	AS	EB	EC	ED	ES	EU	FC	FK	FS	FW	IM	IS	KR	LD	LK	OP	OS	OW
query	4	4	4	4	4	4	3	4	4	4	4	3	4	4	4	4	4	4	4	4
click	3	3	3	3.5	3	2	3	3	3	3	3	3	3	3	3	4	4	4	4	4
source	4	4	4	4	4	4	4	4	4	4	4	4	4	5	4	4	4	4	4	4
page	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
SERP	6.5	6.8	7.0	5.8	7.1	9.2	6.2	6.3	7.1	6.0	6.9	7.4	5.6	6.0	6.8	6.7	5.8	6.9	7.3	6.0
content	12.1	10.2	11.1	10.5	11.3	7.2	10.5	10.7	10.1	9.1	10.3	15.3	9.7	12.2	11.9	9.4	11.3	11.3	11.3	13.7
bookmark	1	1	1	1	1	0	0	1	1	0	0	1	0	1	1	1	1	1	0	1