Differences in the Use of Search Assistance for Tasks of Varying Complexity

Robert Capra, Jaime Arguello, Anita Crescenzi, Emily Vardell

School of Information & Library Science
University of North Carolina at Chapel Hill
Chapel Hill, NC, USA
[rcapra,jarguell,amcc,evardell]@email.unc.edu

ABSTRACT

In this paper, we study how users interact with a search assistance tool while completing tasks of varying complexity. We designed a novel tool referred to as the *search guide* (SG) that displays the search trails (queries issued, results clicked, pages bookmarked) from three previous users who completed the task. We report on a laboratory study with 48 participants that investigates different factors that may influence user interaction with the SG and the effects of the SG on different outcome measures. Participants were asked to find and bookmark pages for four tasks of varying complexity and the SG was made available to half the participants. We collected log data and conducted retrospective stimulated recall interviews to learn about participants' use of the SG. Our results suggest the following trends. First, interaction with the SG was greater for more complex tasks. Second, the *a priori* determinability of the task (i.e., whether the task was perceived to be well-defined) helped predict whether participants gained a bookmark from the SG. Third, participants who interacted with the SG, but did not gain a bookmark, felt less system support than those who gained a bookmark and those who did not interact. Finally, a qualitative analysis of our interviews suggests differences in motivation and benefits from SG use for different levels of task complexity. Our findings extend prior research on search assistance tools and provide insights for the design of systems to help users with complex search tasks.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: Information Storage and Retrieval

Keywords

Search assistance, search trails, search behavior

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
SIGIR'15, August 09 - 13, 2015, Santiago, Chile.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3621-5/15/08 ...$15.00.
DOI: http://dx.doi.org/10.1145/2766462.2767741.

1. INTRODUCTION

Current search engines are effective in helping users complete simple search tasks such as homepage-finding and fact-finding. However, they provide less support in helping users with complex tasks that may involve exploration, analysis, comparison, and evaluation. Prior work has sought to address this limitation by exploring different types of interactive tools to support and assist search engine users. These include tools to help users formulate better queries [9, 15, 18], to communicate system features [17], to assist with note-taking [8], and tools that display the "search trails" followed by other users who completed a related task [20, 23, 26].

Our focus in this work is on search trails. The idea underlying search trails is simple and intuitive—search engine users may benefit from seeing how someone else approached the same or a similar task. To support this, search trails provide an interactive display with information about how another person searched, and may include the queries issued, results clicked, pages viewed, pages bookmarked, and annotations made by the original searcher. Trails can be created manually, or algorithmically from search log or toolbar data.

Prior research on search trails has focused on measuring the information content of search trails [23], understanding the differences between trails generated by domain experts versus novices [26], developing algorithms for predicting search trails for a given search session [19], and evaluating the usefulness of trail-end point pages for different types of search tasks [21]. While this prior work suggests the usefulness of search trails and their feasibility as a form of search assistance, there have been few controlled laboratory studies to directly evaluate their benefits and use. This is the main focus of our work. We investigate how users engage with an interactive search trail, when they use it (i.e., for which types of tasks), and what benefits they report.

We study user interaction with a search assistance tool we refer to as the *search guide* (SG). Our search guide tool displays the search trails from three users who completed the same task. Each trail shows the sequence of queries that were issued, the results that were clicked, and the pages that were bookmarked. We report on a user study with 48 participants. Participants were given search tasks and asked to use a search engine to find and bookmark pages that would help in constructing a response for the task. Access to the SG was a between subjects variable—24 participants had access to the SG and 24 participants used a control system without the SG. To gain insight into factors affecting user interaction with the SG, the study used a concurrent think-aloud protocol followed by a stimulated recall interview.

In this paper, we investigate five main research questions. In our first research question (RQ1), we study the effects of task complexity on user interaction with the SG. Participants completed four tasks of varying levels of *cognitive* complexity, which refers to the amount of learning and cognitive effort required to complete the task [1]. Prior studies found that more cognitively complex tasks are perceived to be more difficult and require more search effort [2, 3, 12, 24]. We suspected that these characteristics may lead people to seek search assistance more frequently for complex tasks.

Our second research question (RQ2) investigates combined factors of the user and the task that may influence user interaction with the SG. For this question, we consider *pre-task* factors such as the participant's level of interest in the task, prior knowledge and search experience in the task domain, and expectations about the task difficulty.

In our third research question (RQ3), we study whether having access to the SG influences outcome measures that reflect the user's experience during the search task. We consider *post-task* measures such as the level of enjoyment, acquired interest and knowledge, satisfaction with the solution and search strategy, experienced difficulty, and perceived level of system support. Seeking search assistance incurs a cost to the user in terms of time and effort. In this respect, outcome measures such as satisfaction and perceived level of system support are likely to depend on whether the user was successful in seeking search assistance. In our fourth research question (RQ4), we investigate whether post-task outcome measures differed among searches where participants: (1) used the SG and gained from it, (2) used the SG but did not gain from it, and (3) did not use the SG at all. In our final research question (RQ5), we examine the effects of task complexity on *why* and *how* users interact (or choose not to interact) with the SG. This question differs from RQ1 in that qualitative data from our stimulated recall interviews was used to characterize SG use and non-use along three dimensions: (1) motivations for use, (2) benefits gained from use, and (3) reasons for non-use.

2. RELATED WORK

This work is informed by three branches of prior research: (1) studies of task complexity and its effects on users' perceptions, behaviors, and outcomes; (2) studies of help-seeking in information retrieval; and (3) studies of search assistance.

Task Complexity. A large body of prior work has focused on characterizing tasks along different dimensions (see Li and Belkin [16]). One such dimension is task complexity, which is an inherent property of the task, and is independent of the task doer [16]. Different characterizations of task complexity have been proposed. Early work by Campbell [7] characterized task complexity in terms of the number of required outcomes, the number of alternative paths to the outcomes, the level of uncertainty regarding the paths, and the degree of interdependence between paths. Byström and Järvelin [6] defined task complexity based on the *a priori* determinability of the task, a measure of the extent to which a searcher can read the task description and deduce the required outcomes, the information needed to produce the outcomes, and the processes associated with finding the required information. Bell and Ruthven [4] defined task complexity in terms of the *a priori* determinability of the information required to complete the task, the search strategy, and the ability to recognize relevant content. Finally, Jansen *et al.* [12] (and later Arguello *et al.* [3, 2] and Wu *et al.* [24]) characterized task complexity in terms of the amount of learning and cognitive effort required to complete the task. To this end, they adopted a taxonomy of learning outcomes proposed by Anderson and Krathwohl [1] for designing educational materials. In this work, we use this *cognitive* view of task complexity. Prior studies have shown that more cognitively complex tasks are associated with higher levels of expected (pre-task) difficulty [24], higher levels of experienced (post-task) difficulty [2, 24], and higher levels of search activity as indicated by measures derived from queries, clicks, bookmarks, and task completion time [2, 3, 12, 24]. We contribute to this body of literature by investigating whether and how task complexity affects use of search assistance.

Help-Seeking in Information Retrieval. Searchers encounter difficulty in different ways and for different reasons. In a large-scale user study, Xie and Cool [25] found that searchers encounter difficulty with seven general processes: (1) getting started, (2) identifying relevant resources, (3) navigating a resource, (4) constructing queries, (5) constraining the search results, (6) recognizing useful information, and (7) monitoring the task process. They also identified factors that give rise to help-seeking situations including: the user's domain knowledge and search experience, properties of the task (e.g., its complexity), and characteristics of the interface and the quality of the search results.

Search assistance tools provide an opportunity to support users who encounter difficulty. However, there are challenges in creating successful assistance tools. Prior work points to several reasons for why users do not use help systems, including the cost of cognitively disengaging from the main task, the fear of unproductive help-seeking, and the refusal to admit defeat [10]. Prior studies also found that users may not notice the help when they are cognitively engaged in the main task [10, 13] or may prefer to attempt the task on their own before seeking assistance [13].

Search Assistance Tools. Search assistance tools are aimed to help users with different aspects of the search process. In this review, we focus on prior work on search trails.

White *et al.* [21] experimented with a search assistant tool called *popular destinations*. Given a query, the search system presented a set of trail-endpoint webpages for trails originating from similar queries. Results from a user study found that popular destinations were better suited for exploratory tasks, while query suggestions were better suited for known-item tasks. For exploratory tasks, popular destinations were associated with improved perceptions about the search experience, the quality of the information found, and the level of system support. In a follow-up study, White and Chandrasekar [22] proposed a modification to the popular destinations tool to help users with difficult known-item tasks. The proposed tool surfaces 'labels' associated with the probable target page. Labels were generated from anchor text, queries with clicks on the target, and social bookmarks. A query-log analysis suggests that surfacing labels might help users find target webpages faster.

White and Huang [23] analyzed search trails captured using a browser toolbar and compared the usefulness of the first trail page, the last page, and the full trail, which includes visited pages in between. Using different heuristics, full trails were associated with more coverage, diversity, novelty, and utility, suggesting that users have something to gain from seeing the full trail versus only the endpoints.

Yuan and White [26] compared the quality of search trails produced by experts and novices in the medical domain. Study participants were explicitly asked to produce search trails to be used by others. Experts produced trails with more relevant pages, more objective information, and a more logical transition from general to specific information.

An important step in using trails to support searchers is to predict which trail to display for a particular search session. Singla *et al.* [19] formulated the task as predicting the best trail in response to an input query-click pair. They developed different trail-finding algorithms and evaluated their performance retrospectively using toolbar data. Results suggest that different algorithms perform well for different metrics based on coverage, diversity, utility, and relevance, and that based on a particular metric, the best-found trail often outperformed the one followed by the actual user.

Finally, Fisher *et al.* [11] evaluated the usefulness of *knowledge maps* constructed by a single user or different users over several iterations. Knowledge maps are different from trails and consist of bookmarked pages that are organized and annotated to convey the *schema* of the solution. Interestingly, knowledge maps that were iterated upon were found to be more useful precisely because the schema of the solution was easier to understand and intrinsically valuable. This suggests that search trails may become more valuable if they can be extended and curated by users over time.

3. METHODS AND MATERIALS

3.1 User Study

A laboratory study with 48 participants was conducted to investigate our five main research questions (RQ1-RQ5). Participants were undergraduate university students (75% female). The study used a concurrent think-aloud protocol with a retrospective stimulated recall interview. Each participant completed four search tasks of varying levels of cognitive complexity (Section 3.3). Participants were asked to use a live search system to find and bookmark webpages that would be useful in constructing a response for the task. The system used the Bing Web Search API to retrieve results from the open web and allowed participants to issue queries, click and view results, navigate away from a landing page, and bookmark pages. Participants were asked to provide a brief justification when bookmarking each page.

Access to the SG was a between-subjects variable—24 participants were given access to the SG and 24 participants were not. Task complexity was a within-subjects variable—participants completed four tasks associated with four different levels of cognitive complexity (remember, understand, analyze, and evaluate) and all four tasks were from the same domain (Section 3.3). Cognitive complexity was rotated across participants using a Latin square.

The study protocol proceeded as follows. First, participants were asked to complete a consent form and a demographic questionnaire. Next, participants were shown a video describing the bookmarking features of the system. For the 24 participants who were given access to the SG, the video contained an additional section describing its basic functionality. Participants were told that the SG conveyed information about how three other searchers completed a similar task. The video described how each of the three search trails or "paths" showed the queries that were issued and the results that were clicked and bookmarked.

Our study protocol involved having participants think-aloud while they searched. Participants were instructed to narrate their searches by describing their thought processes and actions. To familiarize participants with the system and to help them become comfortable with thinking aloud, we asked them to spend a few minutes exploring the system and trying an example task before starting the main tasks.

All four tasks followed the same procedure. First, participants were asked to read the task carefully and then complete a pre-task questionnaire (Section 3.4). Next, participants were directed to the search interface. During the search, participants were gently prompted to continue thinking aloud if they fell silent. Participants were given 12 minutes to complete each task. A pop-up message notified participants when they had three minutes remaining. After completing the task, participants were directed to a post-task questionnaire (Section 3.4). All four search sessions and think-aloud comments were recorded using Morae screen recording software. After completing all four tasks, participants started the retrospective portion of the study.

During the search sessions where participants had access to the SG, the study moderator used Morae Observer to view the participant's search and mark the points where the participant moused or clicked in the SG. These points were later used in the stimulated recall interview (Section 3.5).

3.2 Search Interface and Search Guide

Search Interface. The search interface used in the study is shown in Figure 1. The interface in the experimental and control conditions looked identical except that the SG was only present in the experimental condition. The system allowed participants to issue queries, click results, bookmark pages, delete bookmarks, and (in the experimental condition) to interact with the SG. The search task description (A) was always displayed directly above the query input box (B). Results were returned using the Bing Web Search API, which produces 50 results per query. The top 10 results were displayed directly below the query input box (C) and pagination controls were shown below the results. Participants used a Chrome web browser with four buttons integrated into the browser bookmark bar (D). These buttons allowed participants to: (1) return to the search page, (2) bookmark the current page, (3) show the current set of bookmarks, and (4) terminate the task. Clicking the "bookmark this page" button displayed a pop-up window (not shown) that prompted participants to: "Briefly describe why you are bookmarking this page." Participants could bookmark any page including pages linked directly and indirectly from the SERP or the SG. Clicking the "show bookmarks" button displayed a pop-up window (not shown) that listed the current set of bookmarks, with justifications included. From the bookmark view page, participants could delete a bookmark if desired. In the experimental condition, the search guide was displayed to the right of the search results (E). We used Javascript and AJAX to log all user interactions on the SERP including scroll and mouse-enter events.

Search Guide. As shown in Figure 1(E), the SG displayed three "paths" taken by three different users who completed the same search task. Participants could explore the paths using tabs (Path 1-3). Each path included the list of queries that were issued by another user and, for each query, the sequence of search results that were clicked and bookmarked. Participants could use an accordion control

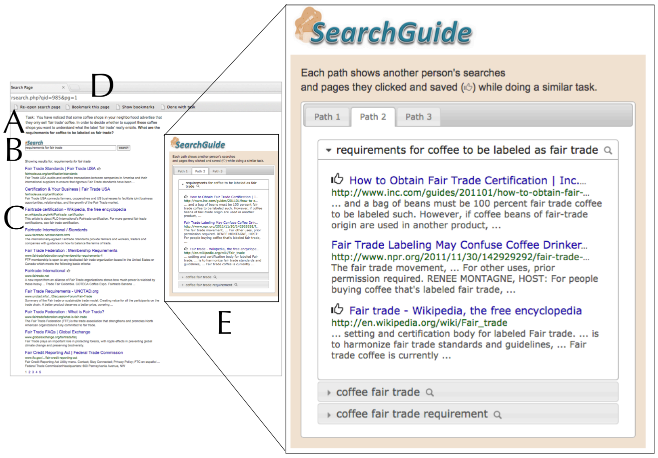


Figure 1: Search Interface and Search Guide

to expand a query to see the sequence of results that were clicked and bookmarked for that query. Clicked and bookmarked pages were displayed using the page title, URL, and summary snippet, and bookmarked pages were distinguished using a thumbs-up icon displayed to the left of the result title. Participants could hover their mouse over a bookmarked page to trigger a tooltip that displayed the justification provided when the page was bookmarked. Clicking on an SG result took the participant to the landing page. Finally, clicking on the magnifying glass icon to the right of an SG query re-issued the query and displayed the results in the main SERP region. Again, we used Javascript and AJAX to record all user interactions with the search guide, including mouse-enter events, clicks on an SG result or query, clicks to expand the accordion control, and tooltip display events.

Two decisions regarding the search guide had to be made—When to display the SG and which paths to display? In regard to the first question, the SG was displayed to participants after issuing the first query and was present on all SERPs for the rest of the search session. In practice, a system might need to predict when to display the SG. We decided against displaying the SG dynamically in order to control how participants experienced the SG and to learn about SG use at *all* points in the search process, including early in the search session. In regard to the second question, as explained in more detail below, we decided to show paths for the *same* search task. In practice, a system might need to predict which paths to show. We were interested in exploring the best-case scenario where the system finds paths that match the user's current search task. However, to avoid biasing participants to use the SG, they were told that the paths corresponded to searches for a *similar* task.

Search Paths. We used a total of 12 search tasks in our study (Section 3.3). Each search task was associated with its own unique set of SG paths. For a given task, participants in the SG condition saw the same SG paths. Paths were selected from a previous user study that included the same search tasks [2]. For each task, we selected three paths that had at least three queries, at least one click per query, and a total of at least three bookmarks.

3.3 Search Tasks

Participants completed four tasks of varying levels of *cognitive* complexity. Cognitive complexity refers to the amount of learning and cognitive effort required to complete the task. We used a subset of 12 tasks from the original 20 tasks developed by Wu *et al.* [24]. The tasks varied across three domains (commerce, health, science) and across four levels of cognitive complexity from Anderson and Krathwohl's Taxonomy of Learning [1]: (1) *remember*: recalling relevant knowledge from long-term memory, (2) *understand*: constructing meaning through summarizing and explaining, (3) *analyze*: breaking material into constituent parts and determining how the parts relate to each other, and (4) *evaluate*: making judgements through checking and critiquing. The tasks were situated in scenarios geared towards our participant population (undergraduate students) [5].

Table 1 shows the four tasks associated with the health domain. Higher-complexity tasks required more information and more mental processing: remember tasks required finding a fact; understand tasks required compiling a list of items; analyze tasks required compiling a list of items and understanding their differences; and evaluate tasks required compiling a list of items, understanding their differences, and making a recommendation.

Remember—You recently watched a documentary about people living with HIV in the United States. You thought the disease was nearly eradicated, and are now curious to know more about the prevalence of the disease. **Specifically, how many people in the US are currently living with HIV?**
Understand — Your nephew is considering trying out for a football team. Most of your relatives are supportive of the idea, but you think the sport is dangerous and are worried about the potential health risks. **Specifically, what are some long-term health risks faced by football players?**
Analyze—Having heard some of the recent reports on risks of natural tanning, it seems like a better idea to sport an artificial tan this summer. **What are some of the different types of artificial tanning methods? What are the health risks associated with each method?**
Evaluate—One of your siblings got a spur of the moment tattoo, and now regrets it. **What are the current available methods for tattoo removal, and how effective are they? Which method do you think is best? Why?**

Table 1: Example Search Tasks from Health Domain

3.4 Pre- and Post-Task Questionnaires

The pre-task questionnaire asked about five measures: (1) level of interest, (2) prior knowledge, (3) prior search experience, (4) *a priori* determinability, and (5) expected difficulty. Questions were asked using five-point scales with labeled endpoints, except level of interest, which used a 7-point scale, and prior search experience, which had 4 choices. We asked participants one question each about their level of interest in the task, prior knowledge about the task, and prior search experience in the task domain. We included three questions about *a priori* determinability. Participants were asked how defined the task was in terms of the (i) expected solution, (ii) the information needed to solve the task, and (iii) the steps required to find the necessary information. These three questions were combined into a single *a priori* determinability scale (Cronbach's α = .777). We included five questions about expected difficulty. Participants were asked about their expected level of difficulty in (i) constructing queries for the task, (ii) understanding the search results, (iii) determining the usefulness of the results, (iv) deciding when to stop gathering information, as well as their (v) expected level of overall difficulty. These five questions were combined into a single expected difficulty scale (Cronbach's α = .848).

The post-task questionnaire asked about nine measures: (1) level of enjoyment, (2) engagement, (3) concentration, (4) acquired interest, (5) acquired knowledge, (6) experienced difficulty, (7) satisfaction, (8) time pressure, and (9) system support. All questions were asked using five-point scales with labeled endpoints. We asked participants one question each about their experienced level of enjoyment, engagement, concentration, and time pressure during the task. Similarly, we asked one question each about their level of acquired interest and knowledge. Consistent with the pre-task questionnaire, we included five questions about experienced difficulty that were combined into a single experienced difficulty scale (Cronbach's α = .853). We asked two questions about satisfaction: (i) satisfaction with the information found and (ii) the satisfaction with the chosen search strategy. These two questions were combined into a single satisfaction scale (Cronbach's α=.792). Finally, we included three questions about system support. Participants were asked whether the system (i) helped them get started, (ii) helped them find resources with useful information, and (iii) provided overall support in completing the task. Again, these three questions were combined into a single system support scale (Cronbach's α=.808).

3.5 Stimulated Recall Interview

After completing the post-task questionnaire for the final search task, a stimulated recall interview was conducted with the 24 participants in the SG condition. For each search task, the study moderator used the Morae markers to identify the first and last use of the SG (if any). For each of these SG uses, the moderator played back a portion of the recording around the point of use and asked a series of structured questions. In order to stimulate the participant's memory of the context, the playback included their think-aloud comments, and started about 10 seconds before the SG use started and continued until the SG use ended. After each playback, the moderator asked questions to elicit: (1) motivations for using the SG and (2) benefits gained from using the SG. At the end of each task, we asked about (3) the times and reasons when the participant purposely avoided using the SG. Participants gave verbal free-form responses to all the questions, which were recorded for later analysis.

Interview Analysis. We used qualitative techniques to analyze participants' responses from the interviews. This analysis involved three rounds of qualitative coding. In the first round, two of the researchers independently coded interviews from four participants (16 interviews) using open coding and then resolved their codes to form an initial set of closed codes for each interview question. In a second round, two researchers used the closed codes on interviews from four additional participants and made refinements to the coding scheme. Then, using the final coding scheme, two researchers each coded half of all interviews and reviewed the codes for the other researcher's half. Any points of disagreement were discussed and resolved by both researchers.

4. RESULTS

In this section we present results from our study. First, we present results of a manipulation check to see whether more cognitively complex tasks were found to be more difficult and required more effort (Section 4.1). Then we present results for each of our main research questions (RQ1-RQ5) in Sections 4.2-4.6, respectively.

4.1 Task Complexity Check

Participants completed four tasks of varying levels of cognitive complexity. As a manipulation check, we first examine whether more complex tasks were found to be more difficult by participants. We focus on four aspects of difficulty: (1) expected difficulty, (2) *a priori* determinability, (3) level of search activity, and (4) experienced difficulty. Expected difficulty and *a priori* determinability were measured using responses to the pre-task questionnaire; level of search activity was measured using behavioral signals captured by the system; and experienced difficulty was measured using responses to the post-task questionnaire. In terms of search activity, we derived behavioral signals from queries, clicks, bookmarks, mouse-overs, scrolls, and elapsed time.

We used ANOVAs to measure the effects of task complexity on all measures. Results are presented in Table 2.[1] Overall, more complex tasks were found to be more difficult in terms of the four aspects of difficulty considered. More complex tasks were associated with higher levels of expected and experienced difficulty, and lower levels of *a priori* determinability. That is, more complex tasks were perceived to be less well-defined in terms of the expected solution, required information, and steps to follow. Finally, more complex tasks were associated with more search activity: more queries, clicks, and bookmarks; lower-ranked clicks and bookmarks; more queries without a click or bookmark; more mouse-enter and scroll events; and required more time to complete. Post-hoc tests found that, in most cases, remember tasks were significantly different from understand, analyze, and evaluate tasks. However, understand, analyze, and evaluate tasks were often indistinguishable. This distinction will be important as we discuss the main results.

4.2 Effect of Task Complexity on SG Use

Our first research question (RQ1) investigates whether task complexity affects user interaction with the search guide. To explore this question, for each of the 96 task sessions where participants had access to the SG (24 participants x 4 tasks), we computed three binary measures that indicate different levels of interaction with the SG:

SGclicked: Represents if the participant clicked somewhere in the SG during the task (1), or not (0). This considered all clicks in the SG, including clicks on the accordion and tab controls to explore the SG queries and paths. This measure is an indication of whether the participant was receptive to search assistance for the task.

SGclickedRQ: Represents whether the participant clicked on at least one SG *result* or *query* during the task (1), or not (0). This measure indicates not only the desire for assistance, but whether the participant found something of interest in the SG.

SGbookmarked: Represents whether the participant bookmarked a page that was discovered by clicking on an SG result (1), or not (0).[2] This measure is an indication of whether the participant gained a direct benefit from interacting with the SG.

Figure 2 shows the number of participants that reached each level of SG interaction, organized by task complexity level (max of 24).

[1]TotalScrollDistance was measured in units equal to the height of the SERP.

[2]Includes bookmarks made directly on an SG result as well as bookmarks found by navigating links from an SG result.

Table 2: **Effects of task complexity on expected and experienced difficulty, *a priori* determinability, and search activity.**

	Remember	Understand	Analyze	Evaluate	F(3,188); p-value	post-hoc
A Priori Det.	4.59 (0.55)	3.90 (0.86)	3.62 (0.82)	3.87 (0.79)	14.14; p=.000	R<U,A,E
Expected Diff.	1.65 (0.64)	2.25 (0.66)	2.50 (0.83)	2.31 (0.73)	12.40; p=.000	R<U,A,E
NumQueries	1.92 (1.44)	3.42 (1.80)	4.81 (3.09)	5.15 (3.09)	17.10; p=.000	R<U<A,E;
NumClicks	3.90 (2.32)	5.71 (2.91)	6.98 (3.66)	7.27 (3.07)	12.36; p=.000	R<U,A,E
ClicksPerQuery	2.43 (1.27)	2.08 (1.69)	1.79 (1.03)	1.78 (0.92)	2.79; p=.042	R<E
AvgClickRank	2.91 (1.36)	4.19 (2.21)	3.62 (1.98)	3.81 (2.18)	3.59; p=.015	R<U
AvgTimeToFirstClick	23.59 (9.10)	26.08 (13.29)	31.34 (20.94)	31.60 (19.45)	2.80; p=.041	-
NumAbandonedQueries	0.38 (0.82)	0.77 (1.02)	1.37 (1.93)	1.25 (1.45)	5.38; p=.001	R<A,E
PctAbandonedQueries	0.11 (0.19)	0.17 (0.19)	0.21 (0.22)	0.20 (0.18)	2.60; p=.054	-
NumBooks	2.02 (1.02)	3.46 (1.24)	4.29 (1.99)	4.67 (2.38)	21.66; p=.000	R<U,A,E;U<E
NumBooksPerQuery	1.37 (0.67)	1.35 (0.98)	1.20 (0.88)	1.10 (0.58)	1.23; p=.301	-
AvgBookRank	2.73 (1.80)	4.03 (2.49)	3.60 (2.33)	3.35 (2.58)	2.62; p=.052	R<U
QueriesWOBooks	0.60 (1.18)	1.23 (1.36)	1.96 (2.26)	1.88 (1.86)	6.50; p=.000	R<A,E
PctQueriesWOBooks	0.17 (0.27)	0.28 (0.25)	0.30 (0.26)	0.31 (0.21)	3.08; p=.029	R<E
NumMouseovers	43.90 (43.25)	79.71 (65.72)	80.42 (64.46)	91.58 (66.13)	5.61; p=.001	R<U,A,E
TotalScrollDistance	1.77 (2.09)	4.81 (5.51)	4.26 (5.27)	5.01 (4.69)	5.07; p=.002	R<U,A,E
TimeToComplete	287.39 (368.46)	461.26 (207.37)	507.50 (194.85)	540.72 (169.04)	9.94; p=.000	R<U,A,E
Experienced Diff.	1.73 (0.90)	2.22 (0.92)	2.24 (0.79)	2.23 (0.80)	4.14; p=.007	R<A,E

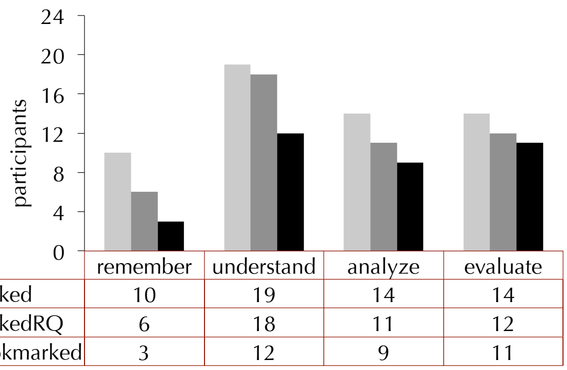


Figure 2: Number of participants (out of 24) who reached each level of SG interaction for different task complexity levels.

As Figure 2 shows, SG interaction was different based on task complexity level. For *SGclicked*, there was a significant effect of task complexity (Cochran's Q test, $Q(3) = 11.372$, $p = .01$). Post-hoc McNemar tests[3] showed that fewer participants interacted with the SG during the remember-level tasks as compared to the understand tasks.

For *SGclickedRQ*, task complexity also had a significant effect ($Q(3) = 15.316$, $p = .002$). Post-hoc tests showed that more participants clicked on SG results and queries during the understand-level tasks as compared to the remember and analyze ones (evaluate was marginally significant).

There was also a significant effect of task complexity on *SGbookmarked* ($Q(3) = 11.038$, $p = .012$). Post-hoc tests showed that fewer participants gained a bookmark from interacting with the SG during the remember tasks as compared to the understand and evaluate ones.

These results show that task complexity had an effect on SG use. As indicated by *SGclicked*, there was an interest in using the SG across all task complexity levels. The effect of task complexity was stronger for measures of interaction that indicate more benefit from the SG use (*SGclicked* and *SGbookmarked*). Finally, the differences in SG interaction were the most pronounced between the remember and understand tasks, with remember tasks having less SG interaction. In Section 4.6, we examine differences in the motivations and benefits that participants described when using the SG during tasks of different complexity levels.

[3]Throughout our analysis, for non-parametric post-hoc tests we use the modified Bonferroni correction outlined by Keppel [14]. For ANOVAs, we use the Tukey correction.

4.3 Effect of Pre-task Factors on SG Use

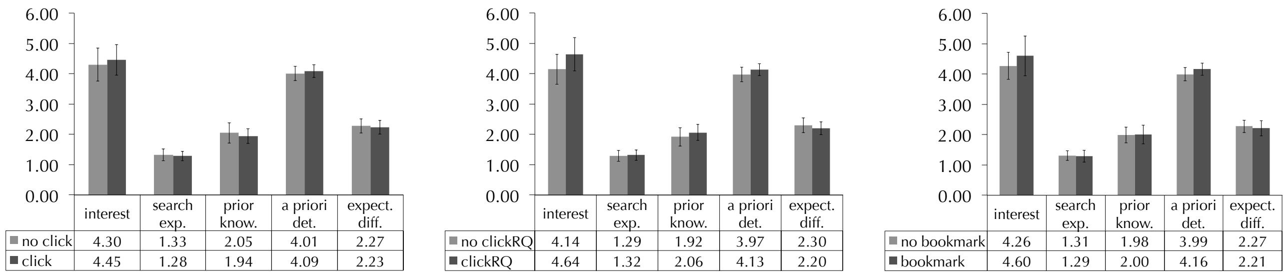
Our second research question (RQ2) investigates whether the factors measured in our pre-task questionnaire (interest, prior knowledge, search experience, *a priori* determinability, and expected difficulty) influenced interaction with the SG. To investigate this question, we ran three logistic regressions to predict the binary measures of SG interaction defined in Section 4.2. Again, this analysis was conducted on the 96 task sessions where participants had access to the SG (24 participants x 4 tasks). Since task complexity was a known source of variance (from RQ1), we also included it as a factor in our model using three indicator variables to distinguish understand (U), analyze (A), and evaluate (E) tasks from remember tasks (treated as the baseline).

Figures 3(a)-3(c) show the mean values of each pre-task factor for each SG interaction measure. For *SGclicked*, the regression model was not statistically significant ($\chi^2(8) = 10.865$, $p = .209$), meaning that the pre-task factors did not significantly predict whether or not a participant clicked on the SG during a task.

For *SGclickedRQ*, the regression model was statistically significant ($\chi^2(8) = 19.907$, $p = .011$). The model explained 25.0% of the variance (Nagelkerke R^2) and correctly classified 66.7% of the cases. Only task complexity was a significant predictor ($p = .01$). *A priori* determinability was marginally significant ($p = .058$).

For *SGbookmarked*, the regression model was statistically significant, ($\chi^2(8) = 17.895$, $p = .022$). Table 3 shows the results. The model explained 23.3% of the variance (Nagelkerke R^2) and correctly classified 70.8% of the cases (an increase of 11% from the 63.5% baseline of always predicting *SGbookmarked* to be zero). Using the Wald criteria, two variables were significant: task complexity ($p = .01$) and *a priori* determinability ($p = .035$). Based on the odds ratio (Exp($\mathcal{B}$)), participants were 2.581 times more likely to gain a bookmark from the SG for every unit increase in the *a priori* determinability of the task.

Overall, results for RQ2 show that none of the pre-task factors were significant predictors of clicked-based interaction with the SG. However, the *a priori* determinability of the task (along with task complexity) was a significant predictor of whether a participant gained a bookmark from the



(a) SGclicked (b) SGclickedRQ (c) SGbookmarked

Figure 3: Mean pre-task factor rating from participants who achieved each level of SG interaction.

SG. In other words, participants were more likely to gain a bookmark from the SG when they perceived the task to be well-defined in terms of the expected solution, required information, and associated steps.

Table 3: Logistic Regression for SGbookmark, $\chi^2(8) = 17.895$, $p = .022$

	$\mathcal{B}$	S.E.	df	p-value	Exp($\mathcal{B}$)
interest	0.14	0.14	1	0.309	1.152
search exp.	-0.31	0.48	1	0.516	0.732
prior know.	-0.12	0.30	1	0.679	0.885
a priori det.	**0.95**	**0.45**	**1**	**0.035**	**2.581**
expect. diff.	0.05	0.44	1	0.914	1.048
complexity			**3**	**0.010**	
complexity(U)	**2.65**	**0.83**	**1**	**0.001**	**14.166**
complexity(A)	**2.31**	**0.87**	**1**	**0.008**	**10.116**
complexity(E)	**2.63**	**0.86**	**1**	**0.002**	**13.869**
constant	**-6.52**	**2.76**	**1**	**0.018**	**0.001**

4.4 Effect of SG Access on Post-task Factors

Our third research question (RQ3) investigates whether access to the search guide had an effect on the factors measured in our post-task questionnaire. To address this, we compare post-task measures of the 24 participants who had access to the SG to the 24 in the control condition.

Figure 4 shows the means and 95% confidence intervals for each post-task measure. We conducted ANOVAs to see if access to the SG influenced the post-task scores on enjoyment, engagement, concentration, interest and knowledge increase, task difficulty, time pressure, satisfaction, and the perceived level of system support. Of these, none were significant except for system support ($F(1, 189) = 10.587$, $p < .001$). Interestingly, participants who had access to the SG reported *lower* levels of system support ($M = 3.80$, $SD = .89$) than participants who did not ($M = 4.27$, $SD = .78$). This result surprised us. Analysis of the effects of SG *use* presented in the next section helps shed light on this result.

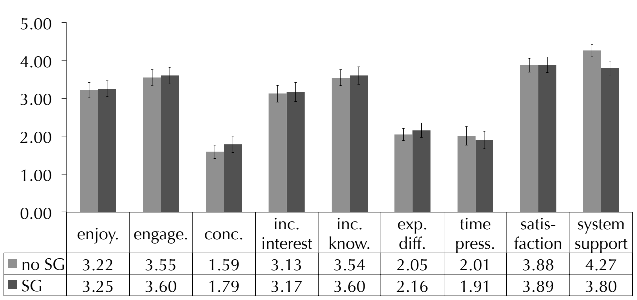


Figure 4: Mean post-task factor ratings the control (no SG) and experimental (SG) group.

4.5 Effect of SG Use on Post-task Factors

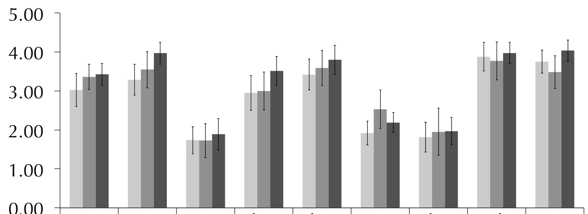
Seeking assistance from the SG required time and cognitive effort from users. For this investment, users may expect to benefit. For RQ4, we investigate whether participants' post-task outcome ratings differed among searches where they used the SG and gained a clear benefit, those where they used the SG but did not gain, and those where they did not use the SG at all. To examine this, we categorized each of the 96 SG task sessions into one of three categories: (1) the participant did not click on the SG ($n = 38$), (2) the participant clicked on the SG, but did not gain a bookmark from using it ($n = 22$), and (3) the participant clicked on the SG and gained a bookmark ($n = 35$).

Figure 5 shows the means for each post-task factor, grouped by category. Results of ANOVAs found a significant effect of category on level of engagement ($F(2, 92) = 3.93$, $p < .023$) and level of experienced difficulty ($F(2, 92) = 3.18$, $p < .046$), and a marginally significant effect on level of system support ($F(2, 92) = 2.76$, $p < .07$). No other post-task factors were significant.

Post-hoc tests showed the following differences. With regard to engagement, when participants gained a bookmark from using the SG, they reported signicantly higher levels of engagement ($M = 3.97$, $SD = .82$, $p = .02$) than when they did not click on the SG at all ($M = 3.29$, $SD = 1.21$). For experienced difficulty, when participants clicked but did not gain, they reported higher levels of experienced difficulty ($M = 2.54$, $SD = 1.12$, $p = .036$) than when they did not click ($M = 1.92$, $SD = .92$). In terms of system support, when participants gained a bookmark, they reported higher levels of system support ($M = 4.04$, $SD = .80$, $p = .058$) than when they clicked but did not gain ($M = 3.48$, $SD = .94$). Interestingly, when participants clicked but did not gain, they reported less system support than when they did not click at all ($M = 3.76$, $SD = .91$), but this difference was not significant ($p = .487$).

Together, the above results suggest that outcome measures that relate to the user experience (e.g., engagement, experienced difficulty, and perceptions of system support) may depend not only on the use of search assistance, but on whether the use is productive and results in a tangible benefit (e.g., a bookmark). These results underscore the importance of providing relevant, high-quality trails.

Finally, these results also provide insight into the surprising result from Section 4.4. Comparing the results for system support in Figures 4 and 5, we see that even participants who gained a bookmark from the SG reported lower levels of system support ($M = 4.04$, $SD = .80$) than the participants in the control condition, who did not have ac-



	enjoy.	engage.	conc.	inc. interest	inc. know.	exp. diff.	time press.	satis-faction	system support
no click	3.03	3.29	1.74	2.95	3.42	1.92	1.82	3.88	3.75
click/no-book	3.36	3.55	1.73	3.00	3.59	2.54	1.95	3.77	3.48
bookmark	3.43	3.97	1.89	3.51	3.80	2.19	1.97	3.97	4.04

Figure 5: Mean post-task factor rating from participants in different SG use groups: (1) did not click, (2) clicked, but did not gain a bookmark, (3) clicked and gained a bookmark.

cess to the SG ($M = 4.27$, $SD = .78$). This suggests that participants in the experimental and control groups had different expectations (or used different grounds for comparison) when responding to our questions about system support. This highlights an important risk in providing search assistance—users' expectations may also increase.

4.6 Differences in SG Use by Task Complexity

Our final research question (RQ5) investigates how SG use and non-use varies for different task complexity levels by evaluating participants' responses during the retrospective stimulated recall interviews. Data from the interviews was used to characterize SG use along three dimensions: (1) motivations for use, (2) benefits from use, and (3) reasons for non-use. We report on participants' free-form responses to our questions using the coding scheme developed.

4.6.1 Motivations for Use

We identified three main categories of motivations for using the SG: (1) to find new information, (2) to confirm previously found information or to confirm the search approach, and (3) to change the search approach. Figure 6(a) shows the number of participants who described each motivation category at least once during the interview (max of 24 for each bar). We elaborate on each category below.

Find new information. Participants described wanting to find new or better information than they had already found, wanting to get closer to an "answer", and wanting to explore what others had found. As Figure 6(a) shows, "find new information" was cited by more participants for the higher complexity level tasks (U=11, A=10, E=12) than for the remember tasks (R=5).

Confirm information already found. As opposed to finding new information, participants also used the SG to confirm information they had already found. This category included motivations to confirm a specific fact, to confirm their search approach, and to confirm the completeness of their findings. This type of motivation was described by more participants for the lower complexity tasks (R=9, U=10) than for the higher complexity tasks (A=4, E=4). This illustrates a theme that we will see again later in this section—for the lower complexity tasks, the SG was used more to confirm information, but for the higher complexity tasks it was used to find new information.

Change approach. Another motivation for using the SG was to help change a participant's search approach. This included using the SG to help get started with the search, to get new ideas for query terms, and to look for divergent or contradictory information. Again this motivation was cited by more participants for the higher complexity levels (U=5, A=5, E=8) than for the remember tasks (R=1).

4.6.2 Benefits Gained from SG Use

We identified four main categories of benefits: (1) gained specific information, (2) gained a new search strategy, (3) reassurance, and (4) no gain. Figure 6(b) shows the number of participants who described each gain category at least once during the interview.

Gained specific information. Participants described a variety of ways that they gained specific information from using the search guide. Responses in this category included: finding relevant web pages in the SG results, directly finding an answer as part of an SG result snippet, finding information that contributed to their knowledge of the task domain, and identifying new dimensions of an answer that they had not considered. Following the same trend as the "find new info" motivation, this benefit was cited by more participants for the higher complexity tasks (U=13, A=9, E=9) than for remember tasks (R=2). Interestingly, the highest number of participants cited this for the understand-level (U=13) tasks, suggesting that there may be characteristics of these tasks that make them well-suited to SG use.

Gained a new search strategy. Participants also described gaining new search strategies through their use of the SG. This category included gaining new query terms to use and getting ideas for search strategies from the paths in the SG. This category followed a similar trend to the "gained specific information" benefit: more participants cited it for the higher complexity tasks (U=7, A=10, E=7) than for the remember tasks (R=1).

Reassurance. Participants described gaining reassurance about a specific source of data, about their search approach, about a specific answer, and reassurance that they had found enough information and not missed anything. Reassurance followed a pattern somewhat opposite to "gained specific info" and "gained a new strategy"; it was mentioned more for the lower-complexity tasks (R=8, U=10) than for the higher ones (A=6, E=3). These results are consistent with the overall trends for the motivation categories—when participants used the SG for lower-complexity tasks, they mainly did so to confirm information they had already found, and the benefits they gained were reassurance that the information they found was good.

No gain. In some instances where the participants interacted with the SG, they reported no gain or benefit. These cases are important to consider because they represent situations where the participant sought help, but the SG failed to provide it. Reports of "no gain" were most prevalent for the lowest (R=5) and highest levels of complexity (E=7), suggesting that the SG was more helpful for tasks at the middle levels of complexity (U=4, A=2).

4.6.3 Non-Use

We identified three main reasons for non-use of the SG: (1) the task was straightforward, (2) the participant preferred to search on their own (at least in the beginning), and (3) reasons related to the novelty and unfamiliarity of the SG. Figure 6(c) shows the number of participants who described each non-use reason at least once during the interview.

Straightforward. One of the most frequent reasons for not using the SG was that the task was straightforward and the participant did not think they needed help for the task. This reason was cited by more participants for the remember tasks (R=12) than for the higher complexity tasks (U=4, A=4, E=4).

Wanted to start the search on their own. In many cases, participants mentioned that they did not use the SG at first, but intended to use it later to verify the quality or completeness of information they found. Participants said that they preferred to start searching on their own more frequently for the higher complexity tasks (U=6, A=8, E=7) than for the remember tasks (R=3).

Novelty/Unfamiliarity. Another reason participants reported for not using the SG was the novelty of the tool or the participant's unfamiliarity with it. We did not notice any trend for this reason across levels of task complexity.

5. DISCUSSION

Our findings provide insights about when, why, and how searchers engaged with search assistance, and about the effects of task complexity on use and benefits of the search guide. Next, we discuss our main findings and implications.

Task complexity influenced help-seeking (RQ1). We found a significant effect of task complexity on user interaction with the SG. Users were more likely to interact with the SG (and to gain a bookmark) for the more complex tasks (understand, analyze, evaluate) than for the least complex (remember). Our manipulation check (Section 4.1) found a similar grouping of task complexity levels with respect to difficulty and search effort. Post-hoc tests showed that remember tasks were consistently different than understand, analyze, and/or evaluate tasks. Together, these findings suggest that more complex tasks were more difficult and led users to seek and benefit more from search assistance. Previously, Xie and Cool [25] suggested that task complexity might influence help-seeking and our findings support this hypothesis. In addition, our results extend work by White *et al.* [21] who found differences in help-seeking behaviors for fact-finding versus exploratory tasks.

Well-defined tasks were more likely to lead to SG interaction and gains (RQ2). In addition to task complexity, the *a priori* determinability of the task also influenced SG interaction and gain. This result suggests that users are more likely to navigate and gain information from someone else's search trail when they perceive the task to be well-defined in terms of the expected solution and the steps required. One possible explanation is that when the task is well-defined, the searcher is better able to understand how another person's search trail may be accessible and beneficial.

An interesting area for future work is to explore ways to improve the accessibility and utility of search trails, especially for less well-defined tasks. In the context of distributed sensemaking applications, Fisher *et al.* [11] found that knowledge maps created by one person were not as easy to understand and as helpful as ones that had been iteratively refined by a sequence of users. Search trails may benefit from a similar approach. Rather than showing the exact trail from a single individual, trails could be iteratively refined and organized. Through this process, the accessibility and usefulness of search trails for more open-ended tasks could be improved by moving the trail toward the most common interpretations and approaches to the task.

Interest, prior knowledge, and search experience did not influence SG use (RQ2). Our results did not find pre-task factors such as level of interest, prior knowledge, and search experience to be significant predictors of SG use. Prior work by Jansen and McNeese [13] considered whether users' self-rated problem solving abilities influenced whether they sought search assistance, but also found no effect. These results suggest that other properties of the system and task (such as task complexity) play a larger role in determining whether assistance is sought.

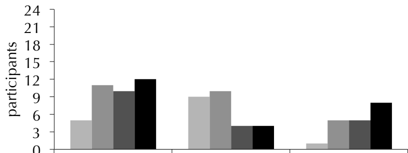
Presenting search assistance can lower impressions of support (RQ3). Participants with access to the SG reported lower levels of system support than participants in the control group. A possible explanation for this is that by adding the search guide, participants' expectations were raised but not fully met, resulting in lower support scores. Another factor may be that having the SG available throughout the task (even at times it was not needed or desired) may have created negative perceptions. This result suggests the importance of showing search assistance dynamically when it is needed. Future work should explore this difference and investigate methods to confidently predict points during the search when assistance is likely to be beneficial.

Users' experience is worse when assistance fails to deliver (RQ4). When our participants interacted with the SG, but did not gain a bookmark from it, they reported experiencing the lowest levels of system support and the highest levels of experienced difficulty. In contrast, when users gained a bookmark from interacting with the SG, they reported the highest levels of system support and engagement within the SG group. These results show that users' perceptions of the search experience can depend on whether or not the use of search assistance was productive. These findings illustrate the importance of predicting the best search trail to display and ensuring that the trail quality is high. Work by Singla *et al.* [19] has reported on techniques to predict the best trails.

Verify and confirm for simple tasks; provide new ideas for complex tasks; allow users to start searches on their own (RQ5). Analysis from our retrospective interviews shows that for the least complex (remember) tasks, when participants used the SG, it was mainly to confirm and verify information. In contrast, for the more complex tasks they used it to find new sources of information and new search strategies. In general, participants did not use the SG to get started, instead preferring to attempt the search on their own before seeking assistance. These results have implications for the design and implementation of search assistance tools. First, since users did not use the SG to get started, search trails do not need to be shown immediately. This represents an opportunity for the system to accumulate evidence about the current task before predicting which search trail(s) to display. Second, the trail selection criteria should consider task type. For simple tasks, the trails could be geared towards verification and confirmation, for example, by showing only the trail endpoints. For complex tasks, the trails could convey more information about the search process or the system could select trails with more divergent information.

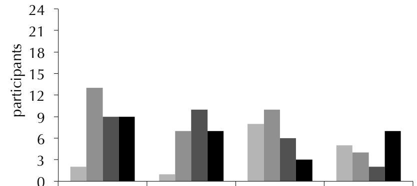
6. CONCLUSION

We reported on a user study that investigated five research questions about user interaction with our search guide (SG) tool. Our findings show that users engaged with the SG and benefited more for complex tasks compared to simpler



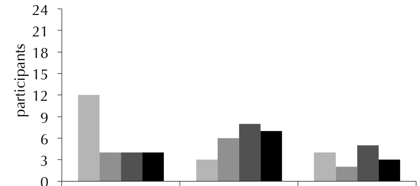
	find new info	confirm info	change approach
R	5	9	1
U	11	10	5
A	10	4	5
E	12	4	8

(a) Motivation for SG use



	specific info	new strategy	reassurance	no gain
R	2	1	8	5
U	13	7	10	4
A	9	10	6	2
E	9	7	3	7

(b) Benefits from SG use



	straight-forward	start self	novelty
R	12	3	4
U	4	6	2
A	4	8	5
E	4	7	3

(c) Reasons for non-use

Figure 6: Differences in SG use for different levels of task complexity.

ones (RQ1). Tasks that were perceived as well-defined were more likely to lead to benefits from using the SG, but other pre-task factors such as level of interest and prior knowledge did not influence SG use or gain (RQ2). Having access to the SG was not found to impact outcome measures such as enjoyment, engagement, and satisfaction (RQ3). However, system support ratings were lower for participants who had access to the SG, suggesting that expectations differed when the SG was shown. When participants interacted with the SG but did not gain a bookmark, they reported higher levels of difficulty and lower levels of system support compared to searches where they did gain or did not use the SG (RQ4).

Analysis of our qualitative results shows that task complexity had an effect on participants' motivations for SG use, benefits from SG use, and reasons for non-use. For the least complex tasks, participants mostly relied on the SG for confirmation and reassurance, and when they did not use it, it was because the task was straightforward. For the more complex tasks, participants relied on the SG to find new information or search strategies, and when they did not use it, it was because they preferred to start on their own.

Our findings point to several directions for future work. Behavioral measures that vary with task complexity may be useful features for predicting when to offer search assistance. Another challenge is how to make search trails more accessible for tasks that are not well-defined and for which users are likely to diverge widely in their approaches. Finally, depending on the task complexity, users are likely to have different motivations for interacting with search trails. Future work might consider customizing the trail display or the trail-finding algorithm to fit different goals (e.g., confirmation vs. finding new information).

7. REFERENCES

[1] L. W. Anderson and D. R. Krathwohl. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman, 2001.

[2] J. Arguello. Predicting search task difficulty. In *ECIR*. Springer-Verlag, 2014.

[3] J. Arguello, W.-C. Wu, D. Kelly, and A. Edwards. Task complexity, vertical display and user interaction in aggregated search. In *SIGIR*, pages 435–444. ACM, 2012.

[4] D. J. Bell and I. Ruthven. Searchers' assessments of task complexity for web searching. In *ECIR*, pages 57–71. Springer-Verlag, 2004.

[5] P. Borlund. Experimental components for the evaluation of interactive information retrieval systems. *Journal of Documentation*, 56(1):71–90, 2000.

[6] K. Byström and K. Järvelin. Task complexity affects information seeking and use. *Inf. Process. Manage.*, 31(2):191–213, 1995.

[7] D. J. Campbell. Task complexity: A review and analysis. *The Academy of Management Review*, 13(1):40–52, 1988.

[8] D. Donato, F. Bonchi, T. Chi, and Y. Maarek. Do you want to take notes?: Identifying research missions in yahoo! search pad. In *WWW*, pages 321–330. ACM, 2010.

[9] H. Duan, Y. Li, C. Zhai, and D. Roth. A discriminative model for query spelling correction with latent structural svm. In *EMNLP-CoNLL*, pages 1511–1521. Association for Computational Linguistics, 2012.

[10] G. Dworman and S. Rosenbaum. Helping users to use help: Improving interaction with help systems. In *CHI*, pages 1717–1718. ACM, 2004.

[11] K. Fisher, S. Counts, and A. Kittur. Distributed sensemaking: Improving sensemaking by leveraging the efforts of previous users. In *CHI*, pages 247–256. ACM, 2012.

[12] B. J. Jansen, D. Booth, and B. Smith. Using the taxonomy of cognitive learning to model online searching. *Inf. Process. Manage.*, 45(6):643–663, Nov. 2009.

[13] B. J. Jansen and M. D. Mcneese. Evaluating the effectiveness of and patterns of interactions with automated searching assistance. *JASIST*, 56:1480–1503, 2005.

[14] G. Keppel and T. D. Wickens. *Design and Analysis: A Researcher's Handbook*. Prentice Hall, 3rd edition, 1991.

[15] Y. Kim and W. B. Croft. Diversifying query suggestions based on query documents. In *SIGIR*, pages 891–894. ACM, 2014.

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

[17] N. Moraveji, D. Russell, J. Bien, and D. Mease. Measuring improvement in user search performance resulting from optimal search tips. In *SIGIR*, pages 355–364. ACM, 2011.

[18] M. Shokouhi. Learning to personalize query auto-completion. In *SIGIR*, pages 103–112. ACM, 2013.

[19] A. Singla, R. White, and J. Huang. Studying trailfinding algorithms for enhanced web search. In *SIGIR*, pages 443–450. ACM, 2010.

[20] A. Wexelblat and P. Maes. Footprints: History-rich tools for information foraging. In *CHI*, pages 270–277. ACM, 1999.

[21] R. W. White, M. Bilenko, and S. Cucerzan. Studying the use of popular destinations to enhance web search interaction. In *SIGIR*, pages 159–166. ACM, 2007.

[22] R. W. White and R. Chandrasekar. Exploring the use of labels to shortcut search trails. In *SIGIR*, pages 811–812. ACM, 2010.

[23] R. W. White and J. Huang. Assessing the scenic route: Measuring the value of search trails in web logs. In *SIGIR*, pages 587–594. ACM, 2010.

[24] W.-C. Wu, D. Kelly, A. Edwards, and J. Arguello. Grannies, tanning beds, tattoos and nascar: evaluation of search tasks with varying levels of cognitive complexity. In *IIIX*, pages 254–257. ACM, 2012.

[25] I. Xie and C. Cool. Understanding help seeking within the context of searching digital libraries. *JASIST*, 60(3):477–494, 2009.

[26] X. Yuan and R. White. Building the trail best traveled: Effects of domain knowledge on web search trailblazing. In *CHI*, pages 1795–1804. ACM, 2012.