Building the Trail Best Traveled: Effects of Domain Knowledge on Web Search Trailblazing

Xiaojun Yuan
University at Albany, State University of New York
Albany, NY 12222
xyuan@albany.edu

Ryen W. White
Microsoft Research
Redmond, WA 98052
ryenw@microsoft.com

ABSTRACT

Web users can help guide others through complex tasks in unfamiliar domains by creating ordered sequences of queries and Web pages, an activity we call *trailblazing*. The trails generated from this process can be surfaced by search engines to help users engaged in these tasks. However, if search engines are going to have people generate trails they need to understand whether there is value in using domain experts for trailblazing (or whether novices are sufficient). In this paper, we describe the findings of a user study of trailblazing in the medical domain, comparing domain novices and experts. We observed differences in how people in each of the groups blazed trails and the value of the trails they generated; experts were more efficient and generated better-quality trails. Although there has been significant research on contrasting novice and expert search behaviors, to our knowledge there is no work (at least in the search domain) on establishing whether artifacts created by domain experts (trails in our case) are more valuable than those created by novices. The answer to this question is important for system designers who want to learn whether investing in domain expertise is worthwhile.

Author Keywords

Trailblazing; domain knowledge; search systems; search trails; search behavior.

ACM Classification Keywords

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval–*search process*; *selection process*.

INTRODUCTION

Web search engines typically return lists of items ranked according to their estimated query relevance. Information retrieval (IR) researchers have worked extensively on algorithms to effectively rank Web documents (c.f. [22]). However, individual items are often insufficient for complex tasks such as understanding medical conditions, planning a vacation, or buying a home [3,19,25].

CHI'12, May 5–10, 2012, Austin, Texas, USA.

Some Web search engines now offer manually-curated lists of sites for particular tasks created by human editors. The Editors' Picks feature of the Microsoft Bing search engine (bing.com/editors-picks) is one example of such functionality. However, when attempting complex tasks, people may need support that extends beyond a ranked list, and actually guides them through the steps required for task completion [16]. Previous work [13,14,17,24,27,28] has shown that trails comprising a filtered set of documents arranged in a useful sequence can help searchers. These trails can provide clear steps for a detailed search scenario and contain information gathered from many different Web domains.

Vannevar Bush [8] envisioned using trails marked and willingly shared by trailblazing users to provide guidance to others as they explored information spaces. He foresaw "a new profession of trail blazers, those who find delight in the task of establishing useful trails through the enormous mass of the common record." Although trails can be generated algorithmically [16,17,24], manual trail generation, or *trailblazing*, more closely aligns with Bush's vision, is unaffected by algorithmic constraints, does not depend on log data, and affords an opportunity for others to directly benefit from skilled users' domain expertise. We anticipate that the adoption of trailblazing tools will become widespread, especially given the recent integration of recommendations sourced from users' social networks into search engines. As such, research in this area is both timely and necessary.

To integrate manually-generated trails into search systems, we must (i) recruit trailblazers with sufficient knowledge to create useful trails, (ii) understand the processes by which they generate trails, and (iii) be able to estimate the value that the generated trails will bring to future searchers (without the need for costly deployment). In this paper, we present the findings of a user study focusing on these three factors and comparing the trailblazing behavior of domain novices and experts. This study was conducted in the context of the medical domain because (i) it is an important area for searchers (seven of ten US adults have searched for medical information online [21]), and (ii) medical searchers frequently have vague or complex needs [9]. As we show, there are differences between novices and experts in the strategies employed to find Web pages for inclusion in the trails, as well as in the trails blazed, including the relevance of the URLs chosen and overall trail usefulness.

This paper makes the following research contributions:

- Proposes the use of trailblazing as a way to leverage domain expertise to help others with less expertise complete complex tasks in the domain of interest.
- Describes the findings of a user study of trailblazing behavior that provides insight into:
  - How users blaze trails in a Web search context.
  - Similarities and differences in how domain experts and novices blaze trails, and if and how the trails eventually generated by the two groups differ.
- Offers design implications for using human-generated trails in search systems, and the role that domain expertise could play in deciding which trails to select.

The remainder of this paper is structured as follows. We present related work. We then describe our user study and its findings. We then conclude by discussing the results and draw from them implications for trailblazing practice, for the design of search systems that leverage human-generated trails, and for automatic trail recommendation algorithms.

RELATED WORK

Several areas of work are relevant to that presented here: (i) modeling information seeking activities beyond basic querying, (ii) mining evidence of trail following behavior from log data, (iii) creating guided tours via human and automatically- generated means, (iv) building computational models of searcher interests to provide step-at-a-time recommendations rather than full trails, and (iv) analyzing the impact of domain expertise on search behavior.

Models of Information Seeking

Models of information seeking have been developed that illustrate the value of navigation well beyond the search result page. O'Day and Jeffries [18] proposed an orienteering analogy to understand users' information-seeking strategies. Their qualitative study relates to ours in describing the benefits of a system that considers the entirety of users' trails. Pirolli and Card [20] developed a sophisticated theoretical model of user behavior known as *information foraging* derived from the patterns exhibited by animals when foraging for food in the wild. The foraging metaphor highlights how information seekers can use cues left by previous visitors to find *patches* of information in a collection, and then consume patch information to satisfy their needs. Continuing work on information foraging theory, Fu and Pirolli [12] developed and validated computational cognitive models of Web navigation behavior based on foraging.

Mining Trails from Logs

Logs containing the search engine interactions of numerous users have been mined extensively to enhance search-result ranking [1,15]. Rich log data, from sources such as browser toolbars, offer insight into user behavior beyond search engine interactions. Search trails comprising query and post-query page views can be mined from these logs [31] and used to help future searchers. White et al. [30] incorporated logged destination pages (terminal trail URLs) corresponding to Web search queries into their search interface prototypes. When presented to user-study participants after the query was submitted, most users found such destination pages useful. Bilenko and White [6] studied full trails mined from logs, including the origin, intermediate, and destination pages. They found that treating the pages in these trails as endorsements improved their ranking in search engines. Finally, White and Huang [34] performed a log-based study to assess search trails followed by users, and showed that the "journey" users took was valuable.

Guided Tours

Guided tours have been proposed as a way to guide users through the steps required to accomplish a task. These tours can be generated manually or automatically.

Manually Generated: Hammond and Allison [14] and Trigg [27] proposed guided tours in hypertext to ease problems of user disorientation. These tours comprised a connected sequence of cards that were presented to users in a predetermined order. Zellweger [37] introduced scripted documents that were more dynamic than guided tours because they included conditional and programmable paths, automated playback, and active entries. Chalmers et al. [10] proposed that human recommenders construct and share Web navigation paths. Wexelblat and Maes [28] introduced annotations in Web browsers called *footprints* that reveal trails through a Web site assembled by the site's designer. Their study found that users required significantly fewer steps to find information using the footprints system.

Automatically Generated: Dispensing with human intervention, tours and trails can also be generated automatically. Guinan and Smeaton [13] generated a tour for a given query based on term-matching for node selection and inter-node relationships (e.g., *is_a*, *precedes*) for node ordering. In a user study based on a collection of lecture materials, they found that users followed these trails closely—40% of the time participants did not deviate from the suggested trail. Wheeldon and Levene [29] proposed an algorithm for generating trails to assist in Web navigation. They defined trails as trees and presented trails to users using a Web browser add-on. Study participants found trails to be useful and noted that seeing the relationship between links helped.

Step-at-a-Time Recommendations

An alternative to presenting the full trail or tour to users is step-at-a-time recommendation. *ScentTrails* [17] combined browsing and searching into one interface by highlighting potentially valuable hyperlinks. Olston and Chi performed user studies with different interfaces incorporating scents of trails within the search results. They showed that users could find information faster and more successfully using the ScentTrails system than by searching or browsing alone. Volant [19] was similar and also highlight hyperlinks of potential user interest during post-query navigation. *WebWatcher* [16] serves as a tour guide agent, accompanying users as they explore the Web. The system highlighted hyperlinks, learned from feedback from earlier tours, that it believed were of interest to the current user.

Domain Expertise

Domain expertise has been studied extensively in the information science community [35]. Studies of domain expertise have highlighted several differences between experts and novices, including: site selection and sequencing [5], task completion time [4], vocabulary and search expression [2], the number and length of queries, and search effectiveness [38]. Bhavnani [4,5] examined domain expert and novice search strategies in the healthcare and shopping domains. Important differences were identified in site selection and knowledge of goal sequencing. Domain experts knew about key resources for their domain and frequently navigated directly to these sites rather than starting with search engines. White et al. [32] used log analysis to study the impact of domain expertise on Web search behavior and found differences in both search behavior (e.g., fraction of queries with technical vocabulary) and resources accessed (e.g., experts focused on technical detail while non-experts focused on consumer-oriented or advisory aspects). These studies focused on differences in the behaviors of domain novices and experts, but did not examine the value of the artifacts they generate. The value is important for system designers since it quantifies one of the primary potential benefits from investing in domain expertise.

Our research extends the work described in this section in a number of ways. First, we focus on the manual generation of trails to assist people in Web search scenarios rather than for particular sites or restricted hypertext corpora. Second, the trails we study are created by humans rather than being mined from search logs or being determined using computational models of search interests such as information scent. Third, we concentrate on blazing full trails, rather than on recommending only next steps, to provide not only guidance and direction, but also an overview of the topic covered. Finally, and most importantly, we study the effect of domain knowledge on trailblazing, and show that experts generate trails more quickly and the trails they generate are of better quality. This suggests that if trails are used by search engines to support users attempting complex search tasks, experts' trails should be preferred.

STUDY

Our user study was designed to obtain a better understanding of the impact of domain knowledge on trailblazing. We describe our research questions, then move onto a description of the tasks, procedure, data capture, and participants.

Research Questions

We were interested in the extent to which domain expertise influences the trailblazing process, as well as the impact of that expertise on the trails that are ultimately created. Specifically, we aim to answer two research questions. First, does domain knowledge affect the trail generation process? Second, does domain knowledge affect the trails that are generated? Among other things, answers to these questions can help search engine companies make sound assessments on the value of recruiting (potentially more costly) domain experts to blaze search trails versus using domain novices.

Tasks

The tasks were designed pursuant to the model of simulated work task situations proposed by Borlund [7]: a scenario description provided background information for a series of activities asked of the participant. All tasks were medical-related, covering the topics of headaches, vertigo, stomach pain, and heart disease. An example task appears in Figure 1. We will return to this task throughout our analysis.

> ***Scenario:*** *You have been bothered by a headache for a while. Your primary doctor suggested that you take a series of tests, but the results revealed nothing serious. One of your friends died of a brain tumor several years ago. You are worried about your own situation and would like to explore the issue in more detail. Specifically, you want to learn more about the types of headache, the corresponding symptoms, the causes, and the remedies for each type. You want to find websites with useful information and share your findings with others by posting as many Web links as you feel are necessary to adequately cover the topic on a social networking site such as Facebook.*
>
> ***Task:*** *Please find as many Web links as you feel are necessary to adequately cover the topic and you believe can be useful to people in a similar situation. Copy and paste the Web links to the answer sheet. In the answer sheet, please answer each question. Also, create a sequence of Web links, comprising all links you selected or some subset, arranged so as to be useful to others.*

Figure 1. Task description for the "headache" task.

The reference to Facebook was included to give participants a readily-understood context for gathering, organizing, and presenting trails of links. Pilot tests were performed to establish whether the tasks were appropriate for the allotted time and if sufficient information was available to construct a trail within the given time frame. Appropriate modifications were made after the pilots to better serve the purposes of the study during the actual experimental phase.

Procedure

We used a between-subjects design. Participants in each group performed four tasks. Task order was randomized based on a Latin-square design to help reduce learning effects. Each study was conducted in person in a laboratory setting and lasted about two hours. When participants arrived, they were welcomed and completed an informed consent form, which included detailed experimental instructions. They then performed the following activities:

1. Completed an entry questionnaire eliciting information about their background, search experience, and experience in organizing information, teaching and creating tutorial information. Inquiring about pedagogical experience was important since trailblazing shares some similarities with tutoring and lesson planning.
2. For each of the four tasks, they:
    a. Completed a pre-task questionnaire;
    b. Conducted a search (per the assigned task) to gather sufficient information to generate a trail. When they felt that satisfactory answers were saved or when they ran out of time (each search was limited to 25 minutes), they continued to the

Variable	Questions
Pre-perceived task difficulty	How difficult do you think it will be for you to find the information for this task
Ease of starting task	Was it easy to get started on this task?
Ease of completing task	Was it easy to complete the task?
Post-task familiarity	How familiar are you with this topic now that you've completed the task?
Enough time	Did you have enough time to complete the task?
Previous knowledge help	Did your previous knowledge of the topic help you?
Satisfaction with found URLs	How satisfied are you with the Web pages you have found?
Confidence in found URLs	How confident are you in the Web pages you collected?
Satisfaction with trail blazed	How satisfied are you with the sequence of Web pages you generated?
Confidence in trail blazed	How confident are you in the sequence of Web pages you generated?
Using the trail blazed	How much do you expect others to learn from the sequence of Web pages that you created?

Table 1. Pre- and post-questionnaire variables.

next topic. The time limitation was imposed in order to have a sufficient number of experimental tasks for statistical validity while keeping the experiment length for each participant manageable;

c. Filled out an answer sheet noting (i) the Web pages accumulated during the search task and (ii) a manually created trail linking the Web pages to be shared with other users (this is the "blazed" trail);

d. Answered a brief post-task questionnaire.

3. After completing all the tasks, participants took part in an exit interview to describe their search experience.
4. Finally, each participant was given 40 USD to thank them for participating in the study.

Data Capture

The Morae usability application (techsmith.com/morae) was used to capture interaction data such as task completion time, number of visited Web pages, queries, and all interaction between each of the participants and the laboratory machine used for the experiment (keystrokes, mouse clicks, etc.). The answer sheet also recorded the sequentially pasted Web pages, the blazed trail (comprising a subset of the recorded pages), and responses to the following three questions about each Web page: (i) *How useful was this page in helping you complete the task*? (ii) *How useful will this page be to people who have had professional medical training*? and (iii) *How useful will this page be to people who have had NO professional medical training?* The answers were expressed according to a seven-point scale (1=*Not at all* to 7=*Extremely*). These ratings are used later in the paper to analyze the utility of the trails blazed by novices and experts.

Measures	Group	Mean	SD	U, p
Familiarity	Expert	3.51	1.81	**U(96)=3422.5 p=0.002**
	Novice	2.73	1.75	
Expertise	Expert	3.03	1.71	**U(96)=3082.5 p<0.001**
	Novice	2.13	1.50	
Difficulty	Expert	3.19	1.52	U(96)=3988.5 p=0.100
	Novice	3.48	1.35	

Table 2. Participant responses to pre-task questions.

The pre-task questionnaire elicits information about participants' level of familiarity with the topic of the task, topic expertise, and perceived pre-task difficulty. The post-task questionnaire focuses on perceptions of the tasks, of the trail generation process for each task, and of the trails generated. Table 1 lists variables and related questions in the pre-task and post-task questionnaires. Only "Pre-perceived task difficulty" comes from the pre-task questionnaire.

In the exit interview, participants were asked to describe their experiences during the experiment and to indicate which factors affected their decisions when choosing Web pages and blazing trails. Interview data were transcribed and analyzed to provide more insight into the trailblazing process. Direct quotations from this analysis are included in the paper to support quantitative analysis as appropriate.

Participants

We recruited 48 participants. Recruitment announcements were sent via email to distribution lists of graduate students at the University at Albany, State University of New York and the Albany Medical Center. Twenty-four graduate students (with no medical training) and twenty-four graduate students majoring in medical-related fields volunteered. Given their medical training, we assumed that the medical graduate students were experts in the medical domain. The non-medical students were deemed to be novices. This assumption was confirmed by analyzing the results of the pre-task questionnaire.

In the pre-task questionnaire, participants rated their topic familiarity, topic expertise, and perceptions of pre-task difficulty on seven-point scales (1=*Not at all* to 7=*Extremely*). The mean (M) and standard deviation (SD) ratings are shown in Table 2, as are the U and p values from the Mann-Whitney U tests. Significance level alpha (α) was set to 0.05. We applied a Bonferroni correction to control for Type I errors (i.e., rejecting true null hypotheses) given multiple comparisons, by setting α to 0.05 divided by the number of dependent variables, in this case 3. Significant differences are bolded in the rightmost column of the table (and in all tables in this paper hereafter).

The results indicate that the medical students were more familiar with the topics and believed that they were more expert. The smaller than expected difference in expertise level between experts and novices may in part be because these assessments are made on a per-task basis, rather than overall across the full medical domain. Although the medical students perceived the tasks to be less difficult than the

Experience	Experts	Novices	U, p
Web search	6.92 (0.28)	6.71 (0.55)	U(24)=239 p=0.118
Medical search	6.13 (1.03)	2.54 (1.82)	**U(24)=37 p<0.001**
Frequency of searching	6.83 (0.38)	6.58 (1.14)	U(24)=282 p=0.849
Information found	6.46 (0.59)	6.38 (0.77)	U(24)=281.5 p=0.880
Search expertise	5.25 (0.61)	5.42 (0.72)	U(24)=251 p=0.395

Table 3. Searching experience of participants. Shown are the mean and (standard deviation) rating values.

Experience	Experts	Novices	U, p
Organizing information	5.13 (1.12)	5.22 (1.00)	U(24)=256 p=0.659
Teaching, tutoring or instructing	5.25 (0.90)	4.74 (1.51)	U(24)=230.5 p=0.221
Creating tutorials or professional presentations	4.71 (1.37)	4.22 (1.70)	U(24)=249 p=0.410
Creating learning outcomes	3.38 (1.28)	3.70 (1.77)	U(24)=238 p=0.289
Structuring course proposals	3.58 (1.47)	3.91 (2.21)	U(24)=242.5 p=0.340

Table 4. Information organizing experience of participants. Shown are the mean and (standard deviation) rating values.

non-medical students, the difference was not significant. This result could be attributed to the simple language used in the tasks, which made them more accessible.

FINDINGS

We divide the presentation of the findings into: (i) *background*: briefly describing relevant details about the participants' skills and experiences, (ii) *process*: detailing the relevant aspects of how participants blazed trails, and (iii) *outcomes*: examining the features of the trails generated.

Background

In the entry questionnaire we gathered information on participants' experiences with Web search and medical search systems. We also gathered information on how successful the participants generally felt when searching and their search frequency. This information helps us better understand what factors are influenced by the novice-expert dichotomy. Table 3 shows the mean and standard deviation ratings on a seven-point scale (1=*Low* to 7=*High*), and significance test results (bold means significant with α=0.01). We can see that experts only significantly differed from novices in terms of medical search experience (which was, as expected, found to be higher for experts).

We also investigated the participants' level of experience in organizing information and in creating tutorial information since we believed that may be important in trailblazing (Table 4). As the table shows, we observed no significant differences between the participant groups in this regard.

Experts	Novices
types of headaches	headache
headache remedies	headache symptoms
headache causes	headache causes
headaches	types of headaches
causes of headache	headache brain tumor
causes of headaches	headache types
migraine remedies	causes of headache
tension headache remedies	brain tumor and headaches
headache types	headaches
headache treatment	

Table 5. Comparison of queries used by experts and novices for task "headache" (user frequency > 2).

Trailblazing Process

We now focus on the first research question and explore aspects of the processes by which domain novices and experts blaze trails. Understanding the process is important not only because Web search is a new application domain for trailblazing, but to help explain observed differences in the trails generated and help develop guidelines for trailblazing practice. Our analysis is mainly from survey and log data, but we augment these findings with data from exit interviews as appropriate. We study a number of aspects of the process: (i) the queries and URLs chosen by participants, (ii) how they constructed the final trail from the URLs recorded on the answer sheet, (iii) the time taken to blaze a trail, (iv) perceptions about the trailblazing process, and (v) the relevance of the URLs selected for the trails.

Searching for Resources

Since previous studies have shown that novices search differently from experts [32], we examined aspects of search behavior among both novices and experts to determine if such differences were also observed during trailblazing.

Queries: Since the queries used may have a large effect on the resources that searchers encounter as they build trails, we begin by analyzing the queries issued by each of the groups. Table 5 presents a frequency-ordered list of the most popular queries used by experts and novices in locating URLs for inclusion in the trails for the "headache" task (those used by at least three participants to reduce noise).

There are a couple of noteworthy differences in the queries from the two groups. First, experts focused on causes and remedies whereas novices focused on understanding the types of headaches, perhaps to improve their knowledge. Second, we observed differences in the conditions that each group targeted. Novices focused on brain tumors, the most concerning aspect of the task description provided to them (see Figure 1). Previous work has shown that the Web is fertile ground for those without medical training to become unduly concerned about serious medical conditions [33]. In contrast, domain experts used queries focused on more common (and benign) explanations for headaches such as tension and migraines. This underscores the importance of domain expertise when searching in a sensitive topic area such as healthcare, and provides a good use case for when trails may have a positive impact on search decisions.

URLs: Previous work has shown differences in the nature of domains that experts and novices access during Web search [32]. We closely reviewed the popular search URLs that the participants encountered as they were searching for resources for the trails. We computed the most popular top-level domains visited by participants, as well as and the distribution of domain name extensions (e.g., .com, .gov) (since White et al. [32] showed differences in this regard).

Both groups used google.com extensively. They also selected significantly similar sites; the popularity ordering of the first six Web sites are the same for the two groups. Experts visited slightly less government and educational sites, and fewer commercial sites. This finding contradicts an analysis of expert and novice behaviors conducted by White et al. [32] who found that medical experts visited more government and educational sites. The difference between the findings of the two studies can be attributed to the intended audience of the trails. Our participants were not blazing trails for themselves, but rather for "others" (from the task description). Experts may have assumed (as intended) that "other" users have little domain knowledge and adjusted their search strategies accordingly. Interview data from experts support this hypothesis: "it would give you the basic information and good overall background information that you would need in order to start your own investigation" and "I think that would help my friends who I don't think are medical professionals to get an idea …"

Assembling the Trails

As stated earlier, participants searched for URLs that might be useful for including in the trails and recorded these URLs on the answer sheet. An important element of the trailblazing process was therefore which URLs transitioned from the answer sheet to the final trail. Experts excluded 49.1% of their gathered URLs, whereas novices excluded 35.8%. Experts appear to be filtering the information they found more, perhaps to better tailor trail content to the target audience. To explore this in more detail, we estimated the reading level of all URLs visited by experts and novices (on a 12-point scale from one to twelve, where each level corresponds to a grade level in the US education system) by using an automatic content-based reading level classifier that uses grade-appropriate vocabulary for each grade [11]. Our findings show that reading level of the documents encountered by experts exceeded that of novices (experts $\underline{M}$=7.54, novices $\underline{M}$=5.33; $U(1814)=324$, $p<0.001$). However, the average reading levels of the trails blazed were similar (experts $\underline{M}$=5.47, novices $\underline{M}$=5.25, no significant difference). This suggests that experts found more sophisticated information, but filtered it for inclusion in the trails.

Time to Blaze a Trail

We measured task completion time from the moment when participants typed in the first query of each task (i.e., once they had read the task description) until they completed the answer sheet. Our findings show that experts ($\underline{M}$=20.05 minutes, $\underline{SD}$=7.75 minutes) spent significantly less time completing their assigned tasks than novices ($\underline{M}$=26.26 minutes, $\underline{SD}$=9.57 mins), $t(190)=-4.936$, $p<0.001$. Experts' familiarity with the domain may have allowed them to more quickly select Web resources and construct the trails.

Perceptions of the Trailblazing Process

In addition to understanding how participants blazed trails, we were also interested in their perceptions of trailblazing task. This may be useful in developing process guidelines, e.g., if we established that novices found it significantly more difficult to start the task then we may want to revise the instructions. This is also useful for understanding additional factors that may affect performance, e.g., if novices felt pushed for time, that may in part explain differences in the trails they generated. These perceptions were collected from the post-task survey and measured on seven-point scales, e.g., 1=*Not at all* to 7=*Extremely*. Table 6 summarizes the responses, averaged over all tasks (bold is significant at α=0.01, corrected for multiple comparisons).

The findings suggest some noteworthy differences, such as it was easier for the experts to complete the tasks, that experts perceived significantly more familiarity with the tasks than did the novices after completing the tasks (although the difference in absolute rating is almost identical to pre-task), and that experts were more comfortable with the time constraints. Also, experts believed that their previous knowledge of the topic helped them significantly more than novices. This concurs with [36], who found that searchers using a domain knowledge visualization system were helped more than those using a generic search system.

Relevance

Finally, we studied the quality of the URLs gathered for the trails. To do so we obtained human relevance judgments for over one hundred thousand queries that were randomly sampled by frequency from the query logs of the Bing commercial search engine. Trained judges assigned relevance labels on a six-point scale—ranging from 1 to 6, and representing *Bad*, *Poor*, *Fair*, *Good*, *Excellent, Perfect*—to top-ranked, pooled Web search results for each query from the Google, Yahoo!, and Bing search engines as part of a separate search engine assessment activity. This provided hundreds of relevance judgments for each query. When intersected with the query-URL pairs present in our data, these judgments allowed us to estimate the relevance of the information that participants in each group added to trails.

Although little overlap existed between the logs of this study and the relevance judgments (115 query-URL pairs in total), significant differences could be found in the average relevance ratings between the expert group and the novice group (experts $\underline{M}$=4.11; novices $\underline{M}$=3.54; $t(113)=2.75$, $p=0.007$). This analysis suggests that, although the Web domains visited by experts and novices were not widely divergent, the URLs that experts selected to include in trails were more query-relevant than those included by novices.

Summary

In this section, we have shown clear differences in the trail generation process between novices and experts. Experts

Measures	Group	Mean	SD	U, p
Ease of starting task	Expert	5.86	1.29	U(96)=3648.5 p=0.010
	Novice	5.38	1.39	
Ease of completing task	Expert	5.43	1.38	**U(96)=3450.0 p=0.002**
	Novice	4.77	1.52	
Post-task familiarity	Expert	5.15	1.20	**U(96)=2861.5 p<0.001**
	Novice	4.30	1.30	
Enough time	Expert	6.14	1.01	**U(96)=3629.5 p=0.008**
	Novice	5.43	1.65	
Previous knowledge help	Expert	3.80	2.01	**U(96)=3329.5 p=0.001**
	Novice	2.83	1.87	

Table 6. Participant perceptions of trailblazing process.

searched for less extreme medical content, spent less time compiling resources and assembling trails, perceived many aspects of the process more favorably, and found more relevant information. We now examine the attributes of the trails that were generated through the trailblazing process.

Trail Analysis

In studying the attributes of the trails generated, we considered (i) the resources they contained, (ii) their structure, (iii) their usefulness, and (iv) participants' perceptions of the resources chosen and trails blazed. All of these attributes can be useful in determining the value of the trails generated, and ultimately can help inform design decisions regarding whether to use experts as trailblazers.

Resources Selected

We examined the features of the URLs selected for inclusion in the trail to determine whether noticeable differences existed in the pages the groups visited. Table 7 shows the most popular top-level Web domains from each group. To better visualize the differences between the lists, we show the rank ordering (in parentheses) of each expert domain in the novice list, and vice versa.

Our earlier analysis showed that novices and experts saved the same top six domains to their answer sheets in the same order. Table 7 shows that the groups selected different resources from those lists for inclusion in the trail. Differences were apparent in the overall popularity of co-occurring URLs, as well as in the URLs selected. There appear to be differences in the types of information that the different groups valued. The most popular domain for novices was google.com, suggesting that they believed search queries were the best way to access domain-specific resources (something also evident in the structure of the trails blazed, described later). In contrast, domain-specific medical sites were more popular than search engines among experts. Interestingly, Bhavnani [4,5] showed similar preferences when comparing novice and expert searching.

Our post-study interviews revealed some similarities in URL selection criteria. Participants preferred information from known, trusted, or renowned Web sites and authored by a health professional. Also, the manner in which the information is presented in the website was important. The

Experts	Novices
1. webmd.com (2)	1. google.com (4)
2. mayoclinic.com (3)	2. webmd.com (1)
3. medicinenet.com (4)	3. mayoclinic.com (2)
4. google.com (1)	4. medicinenet.com (3)
5. en.wikipedia.org (6)	5. emedicinehealth.com (6)
6. emedicinehealth.com (5)	6. en.wikipedia.org (5)
7. ncbi.nlm.nih.gov	7. headaches.org (8)
8. headaches.org (7)	8. nlm.nih.gov
9. cancer.about.com	9. heartdisease.about.com
10. cdc.gov	10. cancer.gov

Table 7. Top-ten most popular domain names selected.

exit interviews revealed that participants in both groups sought out authoritative information (e.g., "a source that's reputable and authoritative and comprehensive in terms of the information", "the best places to look for information are either government or organizations"). As shown earlier, differences in queries issued suggest that even though the their goals were similar, the methods of novices were potentially more unreliable. Interviews also revealed differences in the type of information that the groups selected. Experts selected pages providing summary and structured information (e.g., "it started with a brief overview of each topic of each condition and was structured well") while novices favored content-rich pages (e.g., "if that page encapsulated everything that I had to do for the task").

Trail Structure

We next examined the structure of the trails blazed by experts and novices. Figures 2 and 3 present examples of the trails generated by these groups, represented visually as behavior graphs. These figures are fairly representative of the types of trails that we observed members of each group blazing. The figures reveal a number of structural features of the trails generated. The trail starts with a search engine query, [headaches] in Figure 2 and [headaches symptoms] in Figure 3, proceeds through a sequence of queries and pages, and then terminates. The nodes of the graph represent Web pages that the user has visited: rectangles represent page views and rounded rectangles represent SERPs (queries). The order in which pages were visited is indicated by the arrows in both figures. Since Figure 3 has multiple URL visits directly from the same queries, we also include the sequence order in those pages appear for each query. Among other things, the structure of the trails blazed has direct implications for their presentation to users on SERPs (e.g., longer trails may need to be summarized).

The figures reveal some interesting patterns in the trails that the two groups created. Experts' trails started with background information about the condition and then focused on symptoms and causes. In contrast, novice trails started with the symptoms of the condition and did not provide any background information. They also focused on a single or small set of pages with content that completely satisfied the task, perhaps explaining long queries such as [headaches symptoms causes and remedies]. Novice trails showed more branching from individual queries—multiple

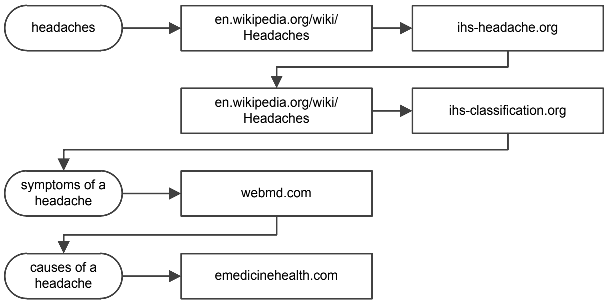

Figure 2. A behavior graph illustrating an expert trail.

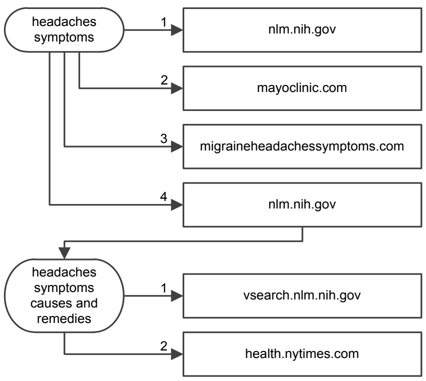

Figure 3. A behavior graph illustrating a novice trail.

sub-trails (i.e., a sequence of URLs beginning with a query and terminating with the end of the session or another query) started from each query. This hub-and-spoke structure perhaps explains why google.com was the most popular domain in trails blazed by novices. In contrast, experts seemed more focused on the ordering of the concepts in the trails. In the exit interviews, experts remarked "I tried to put them logically. Certain things popped up before others but ideally I wouldn't want to read the treatments for stomach cancer before I found out what the tests are to determine if you have it or not" and "… kept it in order of causes and symptoms and then treatments." The strong reliance of novices on queries as hubs is interesting, but also concerning since results may vary significantly over time [26] and new results may appear in the result lists that are not vetted by trailblazers.

In addition to reviewing behavior graphs for the blazed trails, we also computed features of the trails and compared feature values between the experts and novices. Table 8 presents the average feature values of the trails generated by both groups. Given the nature of the data, we use unpaired t-tests, and bold is significant at α=0.008 given the correction for multiple comparisons (i.e., 0.05 divided by 6).

Table 8 shows that novices created significantly shorter sub-trails than experts, perhaps because of the hub-and-spoke strategy that they used. We did not find significant differences in any of the other trail features examined. This result may be because both groups had similarly high searching experience or because trails were being created

Feature		Experts		Novices		t-test
		M	SD	M	SD	
Num. queries		3.70	2.82	2.90	1.95	t(190)=2.29 p=0.023
Length (URLs)		8.53	5.46	10.40	5.31	t(190)=-2.40 p=0.017
Num. sub-trails		3.15	1.84	4.72	3.86	**t(190)=-3.61 p<0.001**
Sub-trail length		4.29	5.51	3.70	3.62	t(190)=0.88 p=0.378
Query length	Token	3.05	0.97	3.14	1.34	t(190)=-0.54 p=0.587
	Chars	21.74	7.11	21.59	8.92	t(190)=0.12 p=0.899

Table 8. Blazed trail feature statistics.

for other users, perhaps leading experts to adjust the length and complexity of queries to suit the target audience.

Trail Value

An important consideration in the decision about whether to invest in domain expertise is the value of the trails blazed. We wanted to establish whether the trails generated for each task were (i) useful in completing the task, (ii) useful for medical experts, and (iii) useful for those without medical expertise. Recall that we asked participants to rate the usefulness of each of the URLs during collection for each of these three groups and record that rating on the answer sheet. However, this only provided insight into a participants own perceptions of the URLs, and we wanted an objective measure of usefulness for each trail, independent of the user who generated it. To do this we computed usefulness scores for each trail URL, averaged across *all users other than the user who generated the trail*. Table 9 shows the mean and standard deviation usefulness values for each group. Ratings are on a seven-point scale (1=*Not at all* to 7=*Extremely*), and bold means significant at α=0.0167 given multiple comparisons.

The results of this analysis show that domain experts created, on average, more useful trails than novices for all three criteria. Since most Web searchers lack formal medical training, it is particularly interesting that experts blazed trails that participants believed would be most useful for searchers without medical expertise. This is critical since helping searchers with low domain expertise perform complex searches is the primary usage scenario for trailblazing.

Participant Perceptions

Finally, we asked participants for their own perceptions of the resources that they found and the trails that they blazed. This gave us a sense for how satisfied and confident users were with their trails, which could be important in making decisions about if and when to use the trails in a practical setting. Table 10 shows the average ratings. We used a seven-point scale (i.e., first four measures, from 1=*Not at all* to 7=*Extremely*; for the last measure, from 1=*None* to 7=*A great deal*), and bold means significant at α=0.01.

The results show that experts were more satisfied and more confident than novices with the trails blazed. In particular,

Measures	Group	M	SD	U, p
Complete task	Expert	5.25	0.66	**U(96)=3162.5 p<0.001**
	Novice	4.84	0.86	
Medical expertise	Expert	4.09	0.80	**U(96)=3304.0 p=0.001**
	Novice	3.68	0.86	
Non-medical expertise	Expert	5.33	0.60	**U(96)=3550.5 p=0.006**
	Novice	5.06	0.78	

Table 9. Participant perceptions of trail usefulness.

Measures	Group	M	SD	U, p
Satisfaction with found URLs	Expert	5.54	1.18	U(96)=3659.5 p=0.011
	Novice	5.02	1.48	
Confidence in found URLs	Expert	5.64	0.91	**U(96)=2997 p<0.001**
	Novice	4.79	1.41	
Satisfaction with trail blazed	Expert	5.56	1.01	**U(96)=3019 p<0.001**
	Novice	4.74	1.44	
Confidence in trail blazed	Expert	5.48	0.98	**U(96)=3298 p<0.001**
	Novice	4.78	1.42	
Using the trail blazed	Expert	5.47	1.04	U(96)=3878.5 p=0.049
	Novice	5.08	1.19	

Table 10. Participant perceptions of the resources found and trails generated.

experts expected others to learn from their trails more than the novices. Although these beliefs do not have any direct bearing on the utility of the trails generated for other users, the heightened satisfaction and confidence from expert users may be a reflection of the higher relevance and utility of the trails they blazed and may be useful as criteria to decide between different experts' trails for the same query or task.

Summary

We have examined the trails generated by experts and novices. Both groups sought authoritative information, but experts preferred overviews, whereas novices preferred content-rich pages. Experts structured their trails as a chain with a clear flow from background to details, whereas novice trails usually had a hub-and-spoke arrangement and focused primarily on details. Expert-generated trails were more valuable and experts also appeared more confident and satisfied with the trails they generated.

DISCUSSION AND IMPLICATIONS

Our analysis showed that experts search differently than novices. Experts pay more attention to the sequence of the task answering; novices focus more on the task content. We also observed that experts rely more on common explanations of the topics while novices concentrated on the most anxiety-producing aspect of the topic, raising concerns about the credibility of the medical URLs in the trails generated (something that has been explored in previous work in this domain [33]). Experts also completed the trailblazing process more quickly even though they had to discard a larger fraction of the URLs that they encountered. In addition, the URLs gathered by experts were more relevant than those visited by novices – giving them a higher quality pool of URLs from which to assemble their trails. The trails blazed by experts were also rated more useful on average.

Our findings suggest that trails can help searchers seek relevant information in a domain of interest, and that trails blazed by experts are likely to help more than those blazed by novices and have the highest potential for significant user benefit. For certain queries or tasks, trails generated by domain experts could be shown to searchers directly to support them in complex search tasks. This is feasible since a large number of trails per task are not required and initially trails could be generated for popular tasks only.

Trails could be displayed on the results page as an alternative to traditional search result lists, as "answers" in addition to result lists, in pop-ups shown after hovering on a result, below each result along with the snippet and URL, or even in-situ on the trail a user is following (as proposed in many systems e.g., [16,17,19]). There are also discovery engines such as StumbleUpon (stumbleupon.com) that could recommend experts' trails or trail components to searchers as they search and surf the Web.

The information gathered during this study can also be used to inform the development of automatic trail generation algorithms e.g., [19,29]. Following experts' example, the algorithms could require that trails begin with background information and become more specific as the trail progresses, be arranged in chains rather than as hubs and spokes, and favor structured overviews over detailed content.

Importantly, given that experts generated better trails in less time, there is a question of whether these benefits justify the additional cost of employing highly-qualified expert trailblazers. Although our findings suggest that domain experts performed best, both in terms of processes and output, the differences between novices and experts are not enormous. Further analysis of the costs and benefits is needed on a case-by-case basis to truly understand whether investment in expert trailblazers is prudent, versus dedicated novice trailblazers or crowdsourced alternatives. More research is also needed to understand how trailblazing generalizes to non-medical domains. Medical professionals are usually trained to communicate complicated medical information to novices. This may not be the case in other domains where such communication is not a primary job function. In those cases experts may not be able to effectively eliminate overly-complicated material.

CONCLUSIONS

We have described a user study on the role of domain expertise in trailblazing behavior on the Web. Our findings demonstrate that domain expertise is important in trailblazing behavior and the trails generated. We have shown differences in the strategies that domain experts employ in blazing trails, including differences in the resources that they select and the structure of the trails that they generate. We demonstrated the value that domain knowledge can bring to trailblazing and showed that domain experts select more relevant and useful trail URLs. Future work involves studying trailblazing in other domains, integrating blazed trails directly into search systems where we will show the trails generated by our experts to novices, and conducting follow-up studies in the laboratory and in the wild to assess the impact of the trails on novices' search performance.

REFERENCES

1. Agichtein, E., Brill, E., and Dumais, S. Improving web search ranking by incorporating user behavior information. *SIGIR*, (2006), 19-26.
2. Allen, B.L. Topic knowledge and online catalog search formulation. *Lib. Quart. 61*, 2, (1991), 188-213.
3. Bates, M.J. The design of browsing and berrypicking techniques for the online search interface. *Online Review 13*, 5, (1989), 407-424.
4. Bhavnani, S.K. Important cognitive components of domain-specific search knowledge. *TREC*, (2001), 571-578.
5. Bhavnani, S.K. Domain-specific search strategies for the effective retrieval of healthcare and shopping information. *SIGCHI*, (2002), 610-611.
6. Bilenko, M. and White, R.W. Mining the search trails of surfing crowds: identifying relevant websites from user activity. *WWW*, (2008), 51-60.
7. Borlund, P. Experimental components for the evaluation of interaction information retrieval systems. *Journal of Documentation 56*, 1, (2000), 71-90.
8. Bush, V. As we may think. *Atlantic Monthly 3*, 2, (1945), 37-46.
9. Cartright, M., White, R.W., and Horvitz, E. Intentions and attention in exploratory health search. *SIGIR*, (2011), 65-74.
10. Chalmers, M., Rodden, K., and Brodbeck, D. The order of things: activity-centered information access. *WWW*, (1998).
11. Collins-Thompson, K. and Callan, J.P. A language modeling approach to predicting reading difficulty. *HLT-NAACL*, (2004), 193-200.
12. Fu, W.-T. and Pirolli, P. SNIF-ACT: a cognitive model of user navigation on the world wide web. *Human-Computer Interaction 22*, 4, (2007), 355-412.
13. Guinan, C. and Smeaton, A.F. Information retrieval from hypertext using dynamically planned guided tours. *ECHT*, (1993), 122-130.
14. Hammond, N. and Allison, L. Travels around a learning support environment: rambling, orienteering, or touring? *SIGCHI*, (1988), 269-273.
15. Joachims, T. Optimizing search engines using clickthrough data. *SIGKDD*, (2002), 133-142.
16. Joachims, T., Freitag, D. and Mitchell, T. WebWatcher: a tour guide for the world wide web. *IJCAI*, (1997), 770-775.
17. Olston, C. and Chi, E. ScentTrails: Integrating browsing and searching on the web. *TOCHI 10*, 3, (2003), 1-21.
18. O'Day, V. and Jeffries, R. Orienteering in an information landscape: how information seekers get from here to there. *INTERCHI*, (1993), 438-445.
19. Pandit, S. and Olston, C. Navigation-aided retrieval. *WWW*, (2007), 391-400.
20. Pirolli, P. and Card, S.K. Information foraging. *Psychological Review 106*, 4, (1999), 643-675.
21. Purcell, K., Rainie, L., Mitchell, A., Rosenstiel, T. and Olmstead, K. Understanding the participatory news consumer. *Pew Internet & American Life Proj.*, (2010).
22. Schwarz, J. and Morris, M.R. Augmenting web pages and search results to support credibility assessment. *SIGCHI*, (2011), 1245-1254.
23. Singhal, A. Modern information retrieval: a brief overview. *Bulletin of the IEEE Computer Society Technical Communication on Data Engineer. 24*, 4, (2001), 35-43.
24. Singla, A., White, R.W., and Huang, J. Studying trailfinding algorithms for enhanced web search. *SIGIR*, (2010), 443-450.
25. Teevan, J., Alvarado, C., Ackerman, M.S., and Karger, D.R. The perfect search engine is not enough: A study of orienteering behavior in directed search. *SIGCHI*, (2004), 415-422.
26. Teevan, J. Adar, E., Jones, R., and Potts, M.A.S. Information re-retrieval: repeat queries in Yahoo's logs. *SIGIR*, (2007), 151-158.
27. Trigg, R.H. Guided tours and tabletops: tools for communicating in a hypertext environment. *TOIS 6*, 4, (1988).
28. Wexelblat, A. and Maes, P. Footprints: history-rich tools for information foraging. *SIGCHI*, (1999), 270-277.
29. Wheeldon, R. and Levene, M. The best trail algorithm for assisted navigation of web sites. *LA-WEB*, (2003).
30. White, R.W., Bilenko, M., and Cucerzan, S. Studying the use of popular destinations to enhance web search interaction. *SIGIR*, (2007), 159-166.
31. White, R.W. and Drucker, S.M. Investigating behavioral variability in web search. *WWW*, (2007), 21-30.
32. White, R.W., Dumais, S.T., and Teevan, J. Characterizing the influence of domain expertise on web search behavior. *WSDM*, (2009), 132-141.
33. White, R.W. and Horvitz, E. Cyberchondria: studies of the escalation of medical concerns in web search. *TOIS 27*, 4, (2009).
34. White, R.W. and Huang, J. Assessing the scenic route: measuring the value of search trails in web logs. *SIGIR*, (2010), 587-594.
35. Wildemuth, B.M. The effects of domain knowledge on search tactic formulation. *JASIST 55*, 3, (2004), 246-258.
36. Yuan, X., Zhang, X.-M., and Trofimvosky, A. (2010). Testing visualization on the use of information systems. *IIiX*, 365-369.
37. Zellweger, P.T. (1989). Scripted documents: a hypermedia path mechanism. *Hypertext*, 1-14.
38. Zhang, X., Anghelescu, H.G.B., and Yuan, X. (2005). Domain knowledge, search behavior, and search effectiveness of engineering and science students. *Information Research 10*, 2, 217.