The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search

Jaime Teevan†, Christine Alvarado†, Mark S. Ackerman‡ and David R. Karger†

† CSAIL, Massachusetts Institute of Technology
Cambridge, MA, USA
{teevan,calvarad,karger}@csail.mit.edu

‡ School of Information, University of Michigan
Ann Arbor, MI, USA
ackerm@umich.edu

Abstract

This paper presents a modified diary study that investigated how people performed personally motivated searches in their email, in their files, and on the Web. Although earlier studies of directed search focused on keyword search, most of the search behavior we observed did not involve keyword search. Instead of jumping directly to their information target using keywords, our participants navigated to their target with small, local steps using their contextual knowledge as a guide, even when they knew exactly what they were looking for in advance. This stepping behavior was especially common for participants with unstructured information organization. The observed advantages of searching by taking small steps include that it allowed users to specify less of their information need and provided a context in which to understand their results. We discuss the implications of such advantages for the design of personal information management tools.

Categories & Subject Descriptors: H.5.2 [**Information Interfaces and Presentation**]: User Interfaces; H.1.2 [**Models and Principles**]: User/Machine Systems – *human factors*

General Terms: Design, Experimentation, Human Factors.

Keywords: Information seeking, search, orienteering, teleporting, context, observational study

INTRODUCTION

Searching for electronic information can be a complex, multistage process, where the information need evolves throughout the course of the search. However, often the search target is known in advance (e.g., a phone number or address). Such small, directed searches have been assumed to be simpler than large, evolving information seeking activities. Researchers have tried to support directed search by attempting to build a "perfect" search engine—i.e., one that returns exactly what is sought given a fully specified information need. Attempts to build such a search engine have focused on improving on keyword search by permitting users to better specify their information need through meta-data [28], natural language [15], and even context [14]. Such a perfect search engine, although perhaps impossible to flawlessly construct, could solve many problems people currently have with search tools.

However, even when a person knows exactly what they are looking for, the perfect search engine might not be enough. Consider Rachel[1], a participant in the observational study presented in this paper. She attempted to locate a document that she knew existed in her file system. Although she knew exactly what document she was looking for (i.e., her information need was not evolving), she could not describe the document, its contents, or its location in advance:

> I don't know how I could have the directory [the document was in] in mind without knowing its name, but I felt sure which it was.

Because she could not specify her information need, a "perfect" search engine probably would not have helped her. Nonetheless, she successfully found her target through a series of small steps, using the local context at each stage of her search to inform her next step.

To understand how to build the best possible search tool, we conducted an observational study of people performing personally motivated searches within their own information spaces. We found that our participants used keyword search in only 39% of their searches, despite almost always knowing their information need up front. Instead of trying to jump directly to their information target using keyword search as might be expected, our participants performed directed situated navigation, similar to the Micronesian islanders' situated navigation described by Suchman [26]. We call this behavior *orienteering* because it is similar to the notion of orienteering in O'Day and Jeffries [21].

[1] All names and identifying details reported have been changed. Minor changes to the transcripts have been made for readability.

In this paper, we explore the range of orienteering behavior we observed, including how it varied between individuals, and discuss its observed advantages, such as that it allowed users to specify less of their information need and provided a context in which to understand any found information. We conclude with a discussion of the implications of our findings for the design of next-generation search tools.

PREVIOUS WORK

Information seeking—where a person's information need evolves throughout the search process—has been well studied. For example, Marchionini [18] detailed the importance of browsing in information seeking and O'Day and Jeffries [21] characterized the information seeking process by outlining common "triggers" and "stop conditions" that guide people's search behaviors as their information needs change. Bates [3] and Belkin [4] proposed search interfaces that allow users to modify and refine their queries as their information need evolves, thus modeling search as an information gathering activity rather than a single, static search.

Directed search has also been well studied, and a variety of methods have been used. Laboratory studies have allowed researchers to conduct controlled studies and examine users' thought processes during search by having them think aloud as they search [7, 16, 19]. However, such studies introduce artificialities that can bias behavior. For example, the search tasks are imposed by the researcher rather than motivated by the user, and task has been shown to affect search performance [12].

To gain a more realistic idea of what search is like in the real world, other studies have examined Web logs. Query log analysis [6, 25] provides insight into the types of information people search for (e.g., sex) and a cursory understanding of how people search (e.g., they use very short queries), but does not provide insight into their underlying intentions. Even when researchers supplement query log analysis with user surveys [6], these studies are still limited to searches that involve search engines, omitting many search activities (61% of the search activities in the study presented here did not involve keyword search). Web site log analysis [11] addresses a broader class of Web behaviors but conflates undirected browsing behaviors and search, whereas we focus solely on search.

Observational studies, such as the one presented here, allow for a deep understanding of naturalistic search behavior. Previous observational studies have focused on users' interaction with various different subsets of their personal information, such as paper documents [13, 17], email [26], files [20], and the Web [10, 24]. Our study is unique in that we focus on directed search and look at behavior across a broad class of electronic types, including email, files, and the Web. By focusing on the communalities of interaction across types, we gain a broader understanding of general search techniques.

METHODS

We conducted 151[2] semi-structured interviews in which 15 participants reported their most recent search activity. We interviewed each participant twice daily on five consecutive days, interrupting them in their offices at unspecified times. We asked them to describe what they had most recently "looked at" and what they had most recent "looked for" in their email, their files, and on the Web. Each semi-structured interview lasted about five minutes. Our method was similar to the diary studies used in many information interaction studies, as well as the Experimental Sampling Method [22]. To supplement these data, we also conducted direct observation and hour-long semi-structured interviews with each participant about their information patterns.

Our participants consisted of 15 graduate students (10 men, 5 women) in Computer Science at MIT. Participants had attended the university from one to seven years; this range allowed us to observe both those in the process of developing their information organization and those with long standing structure. This group is certainly not representative of the general public (e.g., all were expert computer users). However, in our opinion, our participants did reveal some important search issues. In this paper we will extensively discuss the surprising lack of search tool use among this population. Since our participants were familiar with complex information spaces and sophisticated search tools, we believe this lack of tool use is likely to be even more prevalent among the general population.

In the interviews we performed, we used the term "look for" instead of "search". We did this so as not to predispose our participants to think specifically of keyword search. What precisely was meant by "look for" versus "look at" was defined by the participants themselves based on what they considered effort. By allowing participants to self-categorize when they had to exert effort to find information (as in, for example, Bernard [5]), we were able to learn what types of information needs required effort and what techniques were relied on in those cases. We encouraged the participants to give as much detail as possible.

Each short interview was examined independently by the first two authors, and each search incident was coded as to the type of search performed, with an 85% inter-rater reliability. Because the incidents were not randomly selected (e.g., there are temporal patterns in people's information use), we present only qualitatively-based findings here. The data were analyzed using standard qualitative techniques (e.g., [1]). Our findings are exploratory and observational, and as with many qualitatively-based studies, we seek only to analyze interesting phenomena, rather than to confirm existing theory. Accordingly, we present the incidents that emerged as particularly illustrative of the general patterns observed.

---

[2] We inadvertently interviewed one participant 11 times. This participant is labeled "M" in Figure 1.

SEARCH STRATEGIES

We observed that when people searched for specific pieces of information, such as phone numbers or addresses, they generally knew exactly what they were looking for at the onset of their search. We expected our participants to take advantage of this advanced knowledge of their target by using keyword search[3] more often than they would when searching for general information, where the information need often evolves. Surprisingly, only 34 of the 81 searches for specific information that we observed (42%) involved keyword search, compared to 23 of the 42 searches for general information (55%). To understand how our participants performed directed searches, and why they avoided keyword search in many cases, we performed a qualitative examination of our data and uncovered two differing search strategies: orienteering and teleporting.

Orienteering

We observed many directed searches, like the following, where a series of small steps were used to narrow in on the target. Here, although Jim is looking for the office number of a math professor, Connie Monroe, he does not try to find it directly but instead looks for it via her department's page.

> ***Interviewer****:* Have you looked for anything on the Web today?
> ***Jim****:* I had to look for the office number of the Harvard professor.
> ***I****:* So how did you go about doing that?
> ***J****:* I went to the home page of Math Department at Harvard.

Jim went on to explain that he knew there was a specific Web page with her address:

> ***I****:* Did you know it would be there [on a page] or you just hoped it would be there?
> ***J****:* I knew that she had a very small Web page saying, "I'm here at Harvard. Here's my contact information."
> […]
> ***I****:* So you went to the Math department, and then what did you do over there?
> ***J****:* It had a place where you can find people, a link to the page where you can find people and I went to that page and they had a dropdown list of visiting faculty, and so I went to that link and I looked for her name and there it was.

This search by localized or situated navigation is an illustration of what we call *orienteering*. Orienteering involves using both prior and contextual information to narrow in on the actual information target, often in a series of steps, without specifying the entire information need up front. We observed that orienteering was heavily relied upon, even in directed search for specific information. We explore its characteristics further in the following sections.

Teleporting

At the other end of the spectrum from a search strategy that involves many local, situated steps is a strategy we call *teleporting*. When a person attempts to teleport, they try to jump directly to their information target. Teleporting represents the behavior many search engines try to support in their quest to be "perfect". For example, if Jim, instead of browsing to Monroe's office number, had performed a search for, "Connie Monroe, office number," the perfect search engine would have brought him her office number.

As might be expected, we observed incidents of people teleporting (or trying to teleport). For example, to find housing prices in Boston, Alex went to an Internet search engine and entered "real estate prices Boston" into the search box. In doing so, he was trying to jump directly to that information. Of course, we rarely observed perfect teleporting in practice—even in this example, Alex reported having to "browse through all the different graphs and statistics" that the returned site provided. Regardless, we note that participants do sometimes attempt to jump directly to their information target, but also that such attempts were surprisingly rare. In this paper we address why people often chose not to teleport, and what they did instead.

EXPLORING ORIENTEERING

Orienteering denotes a search behavior in which people reach a particular information need through a series of small steps. Within this general class of activities, we observed a range of search behaviors, including variation in the size of the steps taken along the way and the methods chosen to take those steps.

Most commonly, the participant knew definitively how to get into the vicinity of the information in question and made a large step to get to the correct area. Once there, the participant used local exploration to find the information target. As an example, Erica was trying to find a piece of information about Quebec. She first typed the URL "bonjourquebec.com", which she knew to exist, and then she "kept clicking on links from the main page" to get the information she wanted.

Erica's search also illustrates that our participants often associated their information need with a particular information source. Erica associated information on Quebec with the Bonjour Quebec Web site. Our participants made this type of association not only on the Web, but also in their email and files. In another incident, Carla performed a search to determine the location of a meeting. She knew this information was contained within a particular email, so instead of searching for the information (e.g., by doing a keyword search for "Tuesday meeting location"), she searched for the email—the source of the information that she needed.

This ability to associate information with a source was critical in helping participants orienteer to their information target, as participants often remembered a lot about the source. During Carla's search for the email containing the meeting location, she didn't know much about where the meeting was, but once she associated this information with

---

[3] Keyword search includes the Windows file system "Find", the UNIX `grep` or `find` commands, any Web-based search engine, and any keyword search in an email client. The `grep` command allows a user to search for files containing a given word or set of words; `find` allows the user to search for a file by its name.

a particular email she was able to recall a large amount of meta-information to help guide her search, including the folder the email was located in, the date it arrived, who the sender was, and an idea of where it would be visually.

We observed that a person's information target could be associated with a source even when the participant had never seen the target or the source before. This is illustrated in a search Lawrence conducted to determine if a particular company had any job openings. Although he had never been to the company's Web site and did not know the URL, he guessed a URL, typed it in, and successfully reached the company's homepage—the source where he suspected he would find the information he was looking for. There he found a link to a listing of job openings.

These examples of orienteering involved steps made by typing URLs, clicking on links, and navigating through email. We observed a large variety of techniques used to take small steps while orienteering, including keyword search. Carla used keyword search in orienteering when looking to buy an electric toothbrush. She first performed a keyword Web search to find an online pharmacy site. Then, after navigating through the pharmacy site, she performed a site search for electric toothbrushes. Although most of her activity involved keyword search, the strategy she employed was orienteering, taking relatively small steps to narrow in on a goal. As in Bates [2], we wish to draw a distinction between search strategies and tactics: Orienteering and teleporting are strategies; keyword search is a tactic that can be used to achieve either strategy.

Orienteering was not always characterized by a relatively large step followed by local exploration, as in the above incidents. Often it appeared as if the participant was following a path they could not quite articulate but believed to exist. In the following incident, Rachel described navigating down her directory hierarchy using cues at each level to remind her which step to take next:

> ***Rachel***: I didn't know necessarily how to type that path name from memory and so I used the path completion. […] I knew what its name was relative to the directory above it. I didn't know the path down the whole tree.
> ***Interviewer***: Did you ever make any false completions, start with the wrong letter or something?
> ***R***: No.

Compared to the previous incidents, Rachel's steps as she narrowed in on her goal were relatively small. Because her memory of the path and even the target was so vague, these small steps allowed her to reach a target she may not have been able to access using any sort of keyword search.

THE ADVANTAGES OF ORIENTEERING

One could argue that the reason that people rely on orienteering is that the tools available for teleporting do not work well enough yet. For example, one of our participants attempted to teleport but failed. She fruitlessly tried to determine how much to tip hairdressers performing various keyword searches using the words "tip", "hairdresser", "percent", and "gratuities". However, we also observed a number of cases where people chose to orienteer even when teleporting might have worked. For example, Conor had difficulty finding the location of a city in Switzerland. He did not know exactly where to find that information, but he had four map sites bookmarked. Rather than relying on a keyword search directly to locate the city (something many Web search engines explicitly support), he used the bookmarks to access the map sites and then clicked around to see whether he could find a map with the information he was looking for.

This incident with the map site was not an isolated case; we noted many cases where people made no attempt to teleport to their information need, even when teleporting appeared to be a viable option. We believe that orienteering is more than a coping strategy—it appears to hold many advantages even compared to a significantly improved search engine. Here we speculate on three properties of orienteering that appeared, in our data, to be important to our participants: it decreased their cognitive load, allowed them to maintain a sense of location during their search, and gave them a better understanding of their search result. Each of these benefits will have significant design implications.

Cognitive Ease

Orienteering appeared to lessen our participants' cognitive burden during their searches. It did this by saving them from having to articulate exactly what they were looking for and by allowing them to rely on established habits for getting within the vicinity of their information need, thus narrowing the space they needed to explore.

In the incident described in the introduction, Rachel looked for a specific file, but could not articulate the properties or location of that file. She relied on cues during the search process to help her narrow in on the file, saving herself the cognitive burden of specifying the exact file she wanted:

> I knew what directory I thought it would be in. I had this mental idea of which directory it was. It is just that I didn't know necessarily how to type that path name from memory and so I used the path completion to get the directory. […] I didn't know that path down the whole tree. I didn't know how many levels down it was, even though I knew what the name was at the lowest level of that sub-directory.

In a similar situation, Brooks looked for the location of some documentation. She had no sense of where to find the documentation itself, but she did remember that an email she received contained the path to the documentation. Although she did not remember the path to the email either, she recalled meta-information about the email that she could use to help her orienteer to it:

> The last email I read was an email from Bill describing where to find the documentation on [a project]. […] And I looked for it in the research directory which was where I put things that are sort of done for a research. […] I went and tried to look for the email that looked familiar for being the correct one. The only thing I had to go by was that it was probably from Bill. But I wasn't exactly positive on that. And I wasn't sure where it would be anyway.

In the above cases, the participants orienteered because it helped remind them of exactly what they were looking for and how to find it. It would have been difficult for them to describe their search target at the beginning of their search. In other cases, the participants had a good idea of what they were looking for, but had strong associations between their target and an intermediate location. In these cases, orienteering was an automatic response or habit, where the participant used the first route to their target that came into their mind. In the following instance, Fernando orienteered to a paper posted on the Web through a familiar source:

> ***Fernando***: So Web pages, as a result of getting the… lab memo announcement from Tony, I went to [the lab's homepage] and then clicked on publications and then looked at 2001 publications and looked for something to see if it were up there and how it was listed [and so forth].
> ***Interviewer****:* So why did you choose to go that route?
> ***F****:* Because, well I knew it was a [lab] memo and the only thing I know about it was it was with the… Lab and I figured it would be a click or two away from the home page, so I chose to go navigate through the home page and it didn't take me too long to find publications on the lab page. I was just looking at it, it is right there. It is under research, publications.

In other cases, the importance of relying on habit was even more explicit. Here, Brooks had just searched for a restaurant using a path that had been recommended to her, instead of finding it as she normally would:

> ***Interviewer:*** Next time you search for restaurants, how do you think you'll do it?
> ***Brooks:*** Either way. Whichever way I remember first.

These instances suggest that orienteering might sometimes be used because it is easier to recall the intermediate steps than to explicitly name the target.

Sense of Location

The relatively small steps taken in orienteering also appeared to allow participants to maintain a sense of where they were, helping them to feel in control, to know they were traveling in the right direction with the ability to backtrack, and to feel certain they had fully explored the space when unable find what they were looking for. Recent literature suggests that people are bad at understanding the models that search engines use [19], and this finding could suggest why teleporting, in contrast to orienteering, might feel disorienting and untrustworthy to some people.

In a particularly telling incident, Lawrence performed an extensive search to determine if a company was publicly traded. Throughout his search he seemed to try to keep a sense of place. He began by visiting the company's home page via a URL he was emailed and looking at links there:

> I looked at some links on that page… I didn't actually search, I just looked at the headings.

He was unable to locate the information he needed on the company's Web page, so he found another page he thought was relevant in his browser's history. The page in his history was not the homepage for the second site, so he took advantage of the sense of location the URL provided him and deleted a suffix from it to arrive at the site's homepage. Still not finding whether or not the company was public, he went to his browser's homepage by clicking on the home button and attempted to find the information from the financial links located there. When he failed to find the stock price of the company, he felt he had exhaustively explored the space, and concluded that the company must be private, despite not having found an explicit answer.

Although Lawrence's search was complex, involving several different Web sites and much exploration at each one, he explicitly directed the majority of his search in order to stay in a portion of the Web with which he was familiar. Although he began on an unfamiliar Web site, the company's Web site, even his initial step was not blind because he had received an email saying that it existed. He also used the technique of deleting the suffix of a URL to arrive at a site's homepage in order to avoid a blind step to that page, either through the use of a search engine or by guessing the URL. When he failed to find the information on the company's page, he fell back to two sites to which he had been before, at least one of which (his homepage) was very familiar.

Understanding the Answer

Another advantage of orienteering that we observed was that it gave people a context for their results. We saw our participants use the context of the information they found to understand the results and to get a sense of how trustworthy those results were. Context was often essential in helping the participant understand that they had found what they were looking for, as illustrated in the following incidents in which Rachel looked for files:

> ***Rachel***: I listed the directory. I saw it. Let's see, I saw "setup.1.lisp", setup.2.lisp", setup.3.lisp" and "setup.3" was the most recent one. That is the one I took.
>
> ---
>
> ***Rachel***: I was looking for a specific file. But even when I saw its name, I wouldn't have known that that was the file I wanted until I saw all of the other names in the same directory and that made me realize what the naming scheme had been.
> ***Interviewer***: So by itself you wouldn't have known?
> ***R***: By itself, I probably [would have] thought that [it] wasn't right.

The importance of context in understanding search results has been reported previously [9, 15]. We observed orienteering had an added advantage over simply presenting keyword search results with some surrounding context: it allowed participants to arrive at their result along a path they could understand. This process enabled them to understand exactly how the search was performed, and consequently to accept negative results. This understanding is what let Lawrence, in the previously described search, to conclude that the company was not public. The use of context is also illustrated in the following incident in which Alex looked for a particular image but did not find it:

> ***Interviewer:*** So how'd you go about looking around for the bigger logo?
> ***Alex:*** Systematically. I basically clicked on every single button until I was convinced that, in fact, they all used the same style sheet and there was no home for a bigger one… I don't think that it exists on the Web page.

Because Alex controlled the search, he could say that he believed the image couldn't be found.

INDIVIDUAL DIFFERENCES

Thus far, we have described orienteering and its variations in general. We found that some of this variation was due to differences in behavior between individuals. Some individuals relied more on keyword search as a tactic during their search activities than others. Somewhat ironically, these same individuals tended to put more effort into the organization of their information spaces, and thus were better set up to support orienteering.

When we examined people's email use, two groups of individuals emerged: those who found previously received messages in their inboxes most of the time and those who found previously received messages in other email folders. Those who found email in their inboxes almost never spoke of interacting with emails that were not in their inboxes and almost always expected to find messages in their inboxes, implying they did not file their messages in general:

> ***Interviewer***: How did you know [the message] was still in your inbox?
> ***Susan***: I don't know if I'm weird… [but] I don't move stuff.

Those who didn't find email in their inbox almost always went directly to folders and never expected to find messages in their inboxes, implying they regularly filed their messages. For example, Fernando found a previously received email in his inbox during only one of his six email searches. When asked whether or not he had expected to find the message in his inbox he said:

> No… [but then] I thought, "where would I have put this," [and] I hadn't really had a category for that yet, so I kind of still have it in my inbox and I haven't quite decided where I could put it yet.

His response implies that he ordinarily files his messages except in the unusual circumstances where he cannot assign a topic to a message.

Malone [17] classified people as *filers* or *pilers*. Filers organize information using a rigid structure, and pilers maintain an unstructured information organization. Using this classification, we can call those who regularly filed their email *filers*, and those who piled the email they interacted with in their inbox *pilers*. While similar, finer-grained categorizations of people's email behavior have been suggested [27], our study was not designed to explore such distinctions. However, our study does allow for insight into how filers' and pilers' search behaviors varied. We expand on work by Ducheneaut and Bellotti [8] that discussed filers' and pilers' search efficiency by exploring how organization behavior correlates with search behavior.

Participants clustered into filers and pilers on the basis of their email use[4]. Six participants each found email in their inboxes in fewer than 40% of their email search activities (mean=23%), while the remaining seven each found email in their inbox in over 80% of their email search activities (mean=95%). This difference was significant ($p<<0.001$). The difference was not related to one group searching in their email more, as the average number of email searches we observed was not significantly different for the two groups (filers=5.0, pilers=4.4).

Filers and pilers tended to rely on different search tactics when looking for things within their files and on the Web. As can be seen in Figure 1, filers reported performing more searches for files (or information within those files) than pilers (filers=5.0, pilers=2.4, $p<0.02$). A possible explanation for this result might be that filers, despite their organizational efforts, lose information more often. However, we believe a more plausible explanation is based in how filers and pilers interact with their electronic information. Filers are used to assigning specific locations to their electronic information and going to that specific location to retrieve it. Pilers, on the other hand, maintain a more unstructured organization, and commonly must look around even when directly accessing information. For this reason, pilers would be less likely than filers to report some small amount of looking around as a search activity.

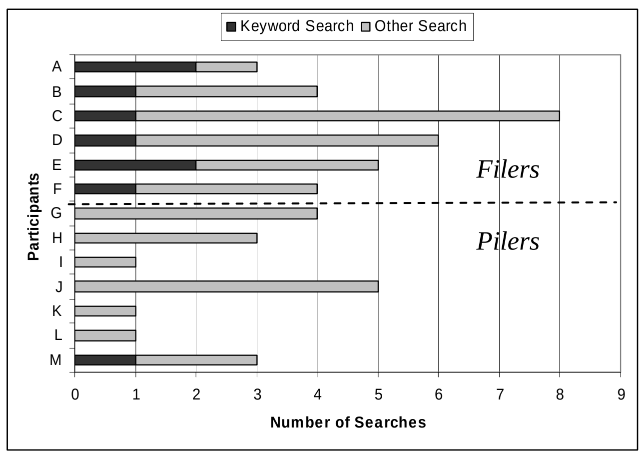

Figure 1: **The number of times participants used each search tactic in their files. Filers searched more than pilers and use keyword search more often.**

We also noted a difference in how often filers and pilers used keyword search to find information. As seen in Figure 1, filers relied more on keyword search when searching in their files than pilers (filers=1.3, pilers=0.1, $p<0.001$). Similarly, while there was not a significant difference in how often filers and pilers used keyword search on the Web (filers=4.0, pilers=5.1), there was a significant difference in the keyword search tools they used. Pilers were more likely to use site search (e.g., search the eBay Web page) as opposed to using a global search engine (e.g., Google). Pilers performed on average 1.7 site searches, while filers only averaged 0.3 site searches ($p<0.02$). Although filers were more likely to use keyword search in their files and global search on the Web, these results do not necessarily

[4] Two participants never reported finding anything in their email, and thus are not included in this discussion.

imply that filers were more likely to teleport to their information target. Keyword search can be used as a tactic in taking small steps toward a target, and we observed both groups orienteering both in their files and on the Web. However, these results do suggest that filers in general tried to take bigger steps when searching for information.

One reason pilers might be more likely to perform site search than filers is that pilers are more used to relying on contextual information to find their information need because they typically navigate through a relatively unstructured information space. Therefore, pilers have developed habits that involve taking more local steps to first arrive at a site before performing a keyword site search to reach their goal. On the other hand, filers are used to assigning meta-data to information in the filing process, and they are more likely to use this meta-data in the retrieval of information through global keyword search. In general, these data suggest that there exist significant individual differences in how people perform directed search. These differences, along with other aspects of orienteering, will be important to consider in the design of new search tools.

DESIGN IMPLICATIONS

In this section we relate what we have learned about people's search behavior to the design of future search tools. Orienteering's prevalence could be due to the fact that search engines do not permit effective teleporting. For example, keyword search engines often fail when confronted with overly specific queries, which could cause people to search first for broad topics and then focus in on their specific target. However, we have seen several advantages to orienteering, including that it appeared to lessen the cognitive burden of finding information, help people better understand their answer, and give people a sense of location during their search. These advantages provide insights for the construction of future search tools that go beyond merely providing perfect keyword search.

To lessen the cognitive burden of search, people used a considerable amount of meta-information during their search that was not available for use by keyword search engines. While search engines are expanding to include meta-data, specifying an information need up front was sometimes more difficult than orienteering to information, and even, in some cases, impossible. A better way of incorporating meta-data is to use meta-data for browsing, as it is in the system being developed by Yee, et al. [28].

We observed that people often looked for a particular information source as a way to find their information target. Searching for the source instead of directly for the target lessened the cognitive burden of search because people often remembered more about the source than they did about the information target itself. Thus, it is particularly important to support the use of meta-data for sources of information, such as Web homepages or email messages. In addition, next generation search tools could learn users' habitually used or trusted sources and make them easily accessible, similar to Maglio and Barrett [16]. Tools could also help people identify the correct source for a given target by previewing the content contained in the source—for example, by automatically flagging email messages that contain email addresses, times, dates or locations.

Orienteering helped participants understand and trust the answers they found. Search tools could enable a similar understanding by showing the context of any results provided (e.g., the source as discussed above, or, in the case of question answering, the context of the answer [15]). Further, search tools could direct search or navigation to sources trusted by the user. To help users understand and accept negative results, search tools could also allow the user to take part in the search process, for example by helping people exhaustively search small areas such as Web pages or individual files.

Orienteering also helped people maintain a sense of location during their search. One technique people used to maintain this sense of location was URL manipulation, which could be better supported by future search tools. In addition, people sometimes knew their target but not the path to that target. To keep users from having to make a blind jump, a next generation search tool could return one or more paths to the potential targets it finds, thus helping the user navigate incrementally. To maintain a sense of location, people often used keyword search engines to take a series of small steps: e.g., first Google, then site search and then page search. A search tool could integrate all three of these variable sized searches into one tool, to keep people from having to think about different tools and different interfaces for each step in their search.

Another way a next-generation search tool could support stepping behavior would be to automatically refine people's information as they interact with it by, for example, clustering the information or suggesting query refinements. Given such a system, the comparison between filers and pilers provides insight into how personalization could be supported. As certain individuals tended to use search engines to take larger steps toward their information targets while others took smaller steps, the size of the refinements could vary according to the size of the step the user is comfortable with. Large, disjoint refinements would be appropriate for users that prefer using keyword search to take large steps, while smaller, similar refinements would be more appropriate for finer-grained navigation.

CONCLUSION

In this paper, we reported on a study of how people looked for information in their email, their file system, and on the Web. We found that often keyword-based search engines were not used when searching, and when they were used, it was usually part of an orienteering strategy. The observed advantages of orienteering include the fact that orienteering allowed participants to not fully specify their information need up front and enabled them to take advantage of the large amount of contextual information they knew about their information target. We have suggested that search tools should support this orienteering behavior.

As with any qualitative study, there are limitations to this study. In this paper, we could only point to the existence of an interesting phenomenon and discuss its implications. Further work will be required to assess how general and pervasive orienteering might be as well as to assess its causal roots and effects. Nonetheless, we believe we have described an important search strategy to support. We acknowledge that our participants were members of the MIT and Computer Science cultures, which highly value information handling. However, in our opinion, the observation that this population, well versed in available search tools, valued orienteering, highlights its importance for dealing with large amounts of electronic information.

We plan to further examine the nature of the contextual information used when orienteering. What people used to search for information appeared to vary based on whether they had seen the information before or not, as well as what type (e.g., email, file, Web) of information being searched for. We will use what we learned to inform the design and development of a next generation information management system [23]. As the amount of information we interact with grows, information management will increasingly become a problem we must deal with. Our study revealed behavioral patterns we can examine further in order to build tools to make this interaction more manageable in the future.

ACKNOWLEDGMENTS

We are grateful to Marti Hearst and Rob Capra for their thoughtful feedback, as well as to the participants in our study for their time. This research was supported by NTT, the Packard Foundation, MIT's Project Oxygen, the Arthur P. Sloan Foundation and the National Science Foundation.

REFERENCES

[1] Ackerman, M. S., and Halverson, C. (1998). Considering an organization's memory. In Proc. of CSCW '98, 39-48.

[2] Bates, M. (1979). Information search tactics. *JASIS,* 30: 205-214.

[3] Bates, M. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review,* 13: 407-424.

[4] Belkin, N. J. (1993). Interaction with texts: Information retrieval as information-seeking behavior. In Proc. of Information Retrieval, 55-66.

[5] Bernard, H. R. (1994). Research methods in anthropology: Qualitative and quantitative approaches. Landham, MD: Altamira Press.

[6] Broder, A. (2002). A Taxonomy of Web search. *SIGIR Forum*, 36(2): 3-10.

[7] Capra, R. G. and Perez-Quinones, M. A. (2003). Re-finding found things: An exploratory study of how users re-find information, Technical Report, Virginia Tech.

[8] Ducheneaut, N. and Bellotti, V. (2001). E-mail as habitat: An exploration of embedded personal information management. *interactions*, 8(5): 30-38.

[9] Dumais, S., Cutrell, E. and Chen, H. (2001). Optimizing search by showing results in context. In Proc. of CHI '01, 277-284.

[10] Jones, W., Dumais, S. and Bruce, H. (2002). Once found, what then?: A study of "keeping" behaviors in the personal use of Web information. In Proc. of ASIST '02, 391-402.

[11] Jul, S. and Furnas, G. W. (1997). Navigation in electronic worlds: Workshop report. *SIGCHI Bulletin*, 29(2):44-49.

[12] Kim, K. S. and Allen, B. (2002). Cognitive and task influences on Web searching. *JASIST*, 52(2): 109-119.

[13] Lansdale, M. (1988). The psychology of personal information management. *Applied Ergonomics*, 19(1): 458-465.

[14] Lawrence, S. (2000). Context in Web search. *IEEE Data Engineering Bulletin*, 23(3): 25-32.

[15] Lin, J., Quan, D., Sinha, V., Bakshi, K., Huynh, D., Katz, B. and Karger, D. R. (2003). The role of context in question answering systems. In Proc. of CHI '03 (Extended Abstracts), 1006-1007.

[16] Maglio, P. P. and Barrett, R. (1997). How to build modeling agents to support Web searchers. In Proc. of UM '97, 5-16.

[17] Malone, T. E. (1983). How do people organize their desks? *ACM Trans. on Office Information Systems*, 1(1): 99-112.

[18] Marchionini, Gary. (1995). Information seeking in electronic environments. New York: Cambridge University Press.

[19] Muramatsu, J. and Pratt, W. (2001). Transparent queries: Investigating users' mental models of search engines. In Proc. of SIGIR '01, 217-224.

[20] Nardi, B. and Barreau, D. (1995). Finding and reminding: file organization from the desktop. *ACM SIGCHI Bulletin*, 27(3): 39-43.

[21] O'Day, V. and Jeffries, R. (1993). Orienteering in an information landscape: How information seekers get from here to there. In Proc. of CHI '93, 438-445.

[22] Palen, L. and Salzman M. (2002). Voice-mail diary studies for naturalistic data capture under mobile conditions. In Proc. of CSCW '02, 87-95.

[23] Quan, D. and Karger, D. R. (2004). Haystack: A personalized, unified interface to creating, browsing, and organizing arbitrary information. In Proc. of CHI '04 (Extended Abstracts).

[24] Sellen, A. J., Murphy, R. and Shaw, K. (2002). How knowledge workers use the Web. In Proc. of CHI '02, 227-234.

[25] Spink, A., Wolfram, D., Jansen, B. J. and Saracevic, T. (2001). Searching the Web: The public and their queries. *Journal of the American Society for Information Science*, 52(3): 226-234.

[26] Suchman, L. A. (1987). Plans and situated actions. New York: Cambridge University Press.

[27] Whittaker, S. and Sidner, C. (1996). Email overload: Exploring personal information management of email. In Proc. of CHI '96, 276-283.

[28] Yee, K-P., Swearingen, K., Li, K. and Hearst, M. (2003). Faceted metadata for image search and browsing. In Proc. of CHI '03, 401-408.