What Are You Looking For? An Eye-tracking Study of Information Usage in Web Search

Edward Cutrell
Microsoft Research
1 Microsoft Way, Redmond, WA 98052
cutrell@microsoft.com

Zhiwei Guan
University of Washington
Box 352195, Seattle, WA 98195-2195
zguan@u.washington.edu

ABSTRACT
Web search services are among the most heavily used applications on the World Wide Web. Perhaps because search is used in such a huge variety of tasks and contexts, the user interface must strike a careful balance to meet all user needs. We describe a study that used eye tracking methodologies to explore the effects of changes in the presentation of search results. We found that adding information to the contextual snippet significantly improved performance for informational tasks but degraded performance for navigational tasks. We discuss possible reasons for this difference and the design implications for better presentation of search results.

Author Keywords
Web search, eye tracking, contextual snippets, user studies.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
As an increasingly large fraction of human knowledge migrates to the World Wide Web and other information systems, finding useful information is simultaneously more important and much more difficult. In 2000, Jansen and Pooch estimated that 1 in 28 of all Web pages that users viewed were search results pages [11]. Today, search is among the most important activities that Web users engage in. Beyond the Web, search is a central activity for users of corporate intranets, specialized databases (from shopping to Medline), and increasingly for personal archives of documents and email [4].

Given the importance and ubiquity of search, it is remarkable how similar most search interfaces are. Users typically type a few words into a query box and receive a rank-ordered list of search results comprising document titles, brief descriptions of the pages and perhaps some metadata about the results (e.g., author, date, size, etc.). On the Web, such interfaces are extremely effective, considering the wide range of tasks they are used for and the very short queries most users provide. However, even given the simplicity of most interfaces, it is not obvious how users utilize different information from lists of search results to complete their tasks. Do users read the descriptions? Are the URLs and other metadata used by anyone other than expert searchers? Does the context of the search or the type of task being supported matter? Eye-tracking methodologies may help us answer such questions by explicitly recording how users attend to different parts of Web search results. As an example, Figure 1 shows users' fixation patterns for a page of Web search results in our study. For this task, users were clearly reading the contextual descriptions, especially on the seventh result.

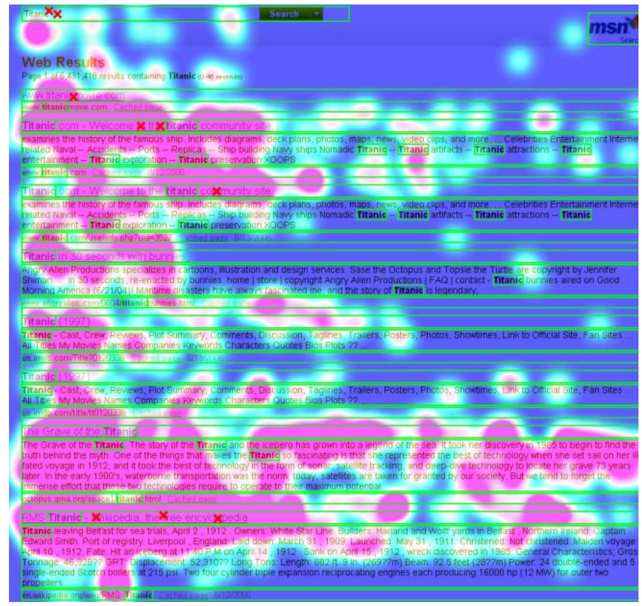

Figure 1. Heat map visualization of the number of fixations across 3 users on a page of search results for an informational task with long contextual snippets (see below). Boxes indicate defined areas of interest (AOIs).

Two or three broad classes of Web search tasks have been identified in the literature [1, 22] and used in various studies on Web search [12, 15, 17]. In *navigational* tasks, users are trying to find a specific Web site or homepage that they have in mind; the goal is simply to get to their destination. In *informational* tasks, the goal is to acquire

CHI 2007, April 28–May 3, 2007, San Jose, California, USA.

some kind of information irrespective of where it might be located. For example, if a user is trying to find out the average June temperature in Caracas, he generally doesn't care where that information is found so long as it is reliable. Earlier research [17] reported that even though informational tasks took longer to complete than navigational tasks, users spent *less* time viewing search result pages for informational tasks (i.e., users spent more time looking for information on destination sites). We thought that this might be because searchers do not have enough information on the search results page to make a good decision about where to go to find what they are looking for.

In this paper we describe an experiment that used eye tracking techniques to help us understand how people use Web search to find information and whether strategies for scanning search results are different for navigational and informational tasks. Do people look at the same number of search results for different task types? Do they attend to different components of search results for navigational and informational tasks? Does the inclusion of more contextual information in search results help with informational tasks?

As noted above, most contemporary Web search engines return a few types of information with each search result: the document title, followed by a short bit of descriptive text, followed by the URL and perhaps some metadata or links (e.g., to cached pages or related links). The descriptive text is usually either a hand-authored description of the page or a *query-dependent contextual snippet*. Snippets are generated on the fly by the search engine from the text in the referenced page based on the query submitted by the user. Snippets typically are 1 to 2 lines of text and often contain sentence fragments from different parts of the document.

Previous work in question-answering suggests that providing additional textual context with answers substantially boosts user preference and performance [16]. In the case of Web searching, additional contextual information poses a trade-off for searchers. On the one hand, even if the longer snippet didn't contain the information itself, we thought that increasing the size of the snippet might help users better decide whether a given result was likely to have what they wanted *before* they navigated to it. However, additional snippet length also could bear substantial costs. First, increasing the snippet length would reduce the number of search results that fit on the screen, forcing users to scroll to see the same number of results. Second, irrelevant search results would include more information, and the "cognitive noise" associated with these snippets potentially could harm performance.

RELATED WORK

Understanding how users search for information on the Web has enormous practical implications for both commercial and academic endeavors. One of the most common techniques for studying Web search is examining search-engine log files [18, 23]. Other researchers have used diary studies to explore the use (and limitations) of search engines in users' daily lives [24]. Researchers at PARC have done careful studies using eye tracking and detailed activity-logging to develop user models for how people explore the Web [3].

As noted above, most Web-search interfaces are extremely simple. While there have been a number of attempts to enrich the UI for search results presentation, relatively few studies have attempted to evaluate how users interact with these new interfaces. Several studies have examined interfaces that organize search results in different ways. "Faceted browsing" interfaces, in which search can be directed through dynamic filtering on orthogonal metadata properties (e.g., for an art database, one might filter on date, artist, media, style, etc.) have been shown to be superior to traditional search interfaces for browsing tasks [25]. Other work has explored interfaces that dynamically categorize search results, grouping similar results together to aid directed search [6, 19]. In Relation Browser++, dynamic categorization and dynamic filtering and visualization of properties were brought together. This interface was significantly better for data exploration tasks than traditional form fill-in interfaces [26].

Eye-tracking methodologies seem particularly promising in the domain of Web search because gaze can be used as a proxy for a user's attention. While many techniques rely on the explicit actions of users (e.g., mouse clicks, query streams or diary reports), eye tracking can yield much more detailed moment-by-moment observations about how users interact with information. Because of this, eye tracking is particularly useful for developing user models (e.g., see [2, 9] for models involving search of computer lists and menus). Joachims, et al. [12] used eye-tracking techniques to characterize how users peruse search results. They then used these observations to inform measures of reliability for implicit feedback from clickthrough data in Web search.

Klöckner, Wirschum and Jameson [13] explored the order in which users examined search results before opening a document. They found that most people employed a linear strategy in which they evaluated each result in turn and decided whether to open that item before moving to the next result. A smaller number (15%) employed a different strategy in which most or all of the results were evaluated before a document was opened.

Lorigo, et al. [17] used measures of fixation, pupil dilation and sequence analysis of patterns of fixations (scanpaths) to look at differences in gender and task type for Web search. For task type, they found that informational tasks took longer to complete than navigational tasks. However, most of the users' time was spent on Web documents and not on search-results pages; users actually spent longer on the search-results pages for navigational tasks. They found no effect of task type on scanpaths (i.e., users evaluated search results in a similar way for both task types). In contrast,

Figure 2. Screenshot of a search-results page from the study. This example includes long query-dependent contextual snippets.

they did find a difference in scanpaths based on gender. Males tended to be more linear in the order in which they looked at results and looked at more results than females.

Understanding how users explore Web search results has large commercial implications as well. A number of companies have emerged that work with businesses in the area of search-engine marketing. These companies help clients develop strategies for increasing traffic to their Web sites (e.g., "search engine optimization"). Detailed understanding of users' behavior and expectations for Web search from eye tracking can be very valuable for these companies and their clients (e.g., [10]).

While most of the work described above investigates Web searching with existing interfaces, none of these studies have examined how users respond to changes in the information provided to them. Joachims, et al. [12] is a possible exception to this because they did manipulate the order of search results for some of their users. The only work we know of that explicitly used eye tracking to explore differences in search interfaces compared a traditional list to a tabular interface for two informational and two navigational search tasks [20]. While no significant differences in performance were found, the eye-tracking measures did turn up a few interesting findings. In particular, they found that the mean number of fixations on the "summary element" (or snippet) for navigational tasks was higher for informational tasks across both interfaces. Unfortunately, this finding may have been driven by the fact that one of their navigational tasks was found to be especially difficult and may have required much more reading for selection confirmation.

EXPERIMENT

To investigate the effect of snippet length on how people use Web search, we designed our study to show results pages in various configurations. First, we presented results with three different snippet lengths (short, medium or long). For another set of questions, we simultaneously varied the position in the search results of the "best" search result for that task. Due to space constraints, we will discuss only the first of these factors, snippet length, in this paper. Analysis and discussion of the manipulation of search-result position is detailed in [8].

In our manipulations, short snippets usually contained a single line of words, medium snippets about two to three lines, and long snippets typically six to seven lines of words. Given our browser and screen size, this meant that when we displayed results with short snippets, seven results were always at least partially visible on the first screen without scrolling. For medium snippets, there were an average of 5.7 (minimum of 5 and maximum of 7), and for long snippets, an average of 4.2 (minimum of 3 and maximum of 6) results were visible on the first screen. The screenshot in Figure 2 shows an example of a query with long snippets, and Table 1 shows the three snippets generated for a single search-result entry. By default, MSN Search presents results with medium length query-dependent snippets. The short and long lengths were chosen to be realistic, but obviously different from the default lengths provided by MSN Search.

All manipulations were performed for two task types: *navigational* and *informational*. In our study, navigational tasks required the participant to find a specific Web page, and informational tasks required the user to find specific information that could be found in one or more places. In

Table 1: Example snippets of each length used in experiment for a single search result.

Welcome to the Oklahoma City Zoo http://www.cpb.ouhsc.edu/OKC/OKCZoo/
Short
The oldest **zoo** in the Southwest and one of the top in the nation, the **Oklahoma** ...
Medium
The oldest **zoo** in the Southwest and one of the top in the nation, the **Oklahoma City Zoo**'s 110 acres are home to more than 2,800 of the world's most exotic animals.
Long
The oldest **zoo** in the Southwest and one of the top in the nation, the **Oklahoma City Zoo**'s 110 acres are home to more than 2,800 of the world's most exotic animals." The Cat Forest/Lion Overlook was completed in 1997. New in 1993 was the Great EscApe , a simulated tropical forest with gorillas, orangutans and chimpanzees. Also found at the **zoo** are the Noble Aquatic Center: Aquaticus , a Children's **Zoo** and Discovery Area, Herpetarium, Island Life Exhibit, Dan Moran Aviary and the Safari Tram. Open 9-5 (Oct-March), 9-6 (April-Sept). Rides additional (weather permitting and seasonal). 2101 N.E. 50th Street **Oklahoma City**, OK (405) 424-3344 (OKC **Zoo** Phone Directory

practice, tasks can vary widely in difficulty and, as seen in [20], this can have a major effect on performance. Therefore it was important to use several different tasks for each task type and to counterbalance across tasks for all manipulations of interest.

Methods

Apparatus

All Web search queries were submitted to a special server for MSN Search (http://search.msn.com) that allowed us to control the length of the snippet information generated for each result. This server used the same methods for dynamic generation of snippets used in production servers (i.e., this used the same set of complex heuristics and algorithms used in MSN Search results). Search results were then dynamically intercepted by a proxy before being rendered to the browser. Advertising and editor-selected content at the top and side of the results page was removed, leaving only the standard UI associated with MSN Search and a list of 10 search results (see Figure 2). Note that because snippets were dynamically generated by the search engine based on indexed content, there were occasions when the snippet for a given item was considerably shorter than those of its neighbors.

Eye tracking was performed using the Tobii x50 eye-tracker (see http://www.tobii.se/) paired with a 17" LCD monitor (96 dpi) set at a resolution of 1024x768. The eye-tracker sampled the position of users' eyes at the rate of 50Hz. An integrated log of eye-movement data, user events and Web pages visited allowed us to map eye movements to various features on the screen during task performance. Areas of interest (AOIs) were generated by a javascript application that parsed the DOM for each page of search results that a user visited. This application provided us with the screen coordinates for each element that we were interested in for a given page (dynamically generated AOIs can be seen in Figure 1).

Participants

Twenty-two participants ranging in age from 18 to 50 years old with a diverse range of jobs, backgrounds and education levels were recruited for this study from a user-study pool. Of these, 4 participants were excluded from these analyses because of stability problems with the eye tracking and/or incomplete data, leaving us with 18 participants (11 male). All participants were moderately experienced at Web search, reporting that they searched the Web for information at least once a week, and all were familiar with several different search engines. None of them had experience using an eye-tracker.

Experimental design and procedure

The design of the experiment crossed *Task Type* x *Snippet Length* and *Task Type* x *Target Position* as independent within-subjects designs. As noted above *Task Type* x *Target Position* is separately analyzed in [8]. For each participant, we randomly varied the order of the search tasks. Each of 12 search tasks (6 different tasks of each type) was counterbalanced across participants such that every task was seen with every snippet length (12x3=36 combinations) and every task was seen at every target position (12x6=72 combinations). We needed to counterbalance across tasks rather than just task types because of the large variability in individual tasks; because some tasks were easier than others, we needed to be able to average across all 6 tasks for each factor we are interested in. For a detailed discussion of the experimental design, please see [6].

Table 2: Search tasks (*queries*) used in study.

Navigational
* Find the homepage of the "Pinewood" software company. (*Pinewood*)
* Find the homepage of the World Cup 2006 soccer games. (*World cup 2006 games*)
* Find the homepage of Comfort Inn. (*Comfort Inn*)
* Find the homepage of the National Weather Center. (*national weather center*)
* Find the homepage of the St. John's law school. (*St Johns Law School*)
* Find the homepage of the Yahoo! People Search. (*Yahoo People search*)
Informational
* Find when the Titanic set sail for its only voyage and what port it left from. (*Titanic)*
* Find the address for the Newark Airport. (*Newark airport address*)
* Find out how long the Las Vegas monorail is. (*Las Vegas monorail*)
* Find out the name of the building that is Piano's most famous work. (*Renzo Piano*)
* Find out the size (in area) of the Oklahoma City Zoo. (*Oklahoma City Zoo*)
* Find the contact number for the Sylvan Learning Center. (*Sylvan Learning Center*)

We designed the search tasks for this study to be representative of common search tasks on the Web, varying in difficulty and topic. On a control page, we gave participants a brief query description and motivation for each task (e.g., *Renzo Piano is a famous architect. Find out the name of the building that is Piano's most famous work*) paired with a link comprising one or more query words that would launch a search when clicked (e.g., *Renzo Piano*).[1] See Table 2 for the list of all tasks and initial queries. After launching the initial query, participants were free to use the search engine however they chose to complete the task. Although participants generally agreed that the initial queries were reasonable for each task, they frequently submitted new queries if they felt they could not find what they were looking for with the query we provided.

[1] We provided the link and query terms for the initial query for every task because we wanted to control the first set of search results that every participant saw. While this does potentially threaten the ecological validity of our experiment, there is considerable benefit in making sure that all users see the same set of initial results.

All the results pages for the initial queries (those generated by the links) were cached locally to ensure that all participants in a given condition would see exactly the same information at the beginning. All search-results pages for subsequent queries were generated on the fly as described above.

For each query we generated, we made sure that the task could be completed with a site presented in the initial set of 10 results. For navigational queries, only one result was associated with the target, while for informational queries there was always one "best" result in which a user could quickly find the searched information (e.g., the searched information was included in the snippet, or the information appeared in a very obvious place on the Web page). However, as is common for informational queries, the task often could be completed by navigating to several different locations if the participant was willing to spend some time "orienteering" around target sites (see [24]).

At the beginning of each session, participants were calibrated for the eye- tracker and given a practice query to familiarize them with the procedure. At the beginning of each task, participants read the task description and motivation in their Web browser and clicked the underlined query when they were ready. Each task was considered completed when participants clicked on the link to the target page, confirmed it was the desired site and vocally announced that they had found the Web site or information requested. At the end of the study, participants provided some demographic information and answered a short questionnaire about their history using Web search engines and their experiences in the study.

Results

Common eye-tracking measures include pupil dilation, fixation information, and sequence information such as scan paths. For our analyses, we relied on measures related to gaze fixations with a minimum threshold of 100 ms in areas of interest. Here we consider AOIs including each individual search result and each sub-element therein (e.g., title, contextual snippet, and URL).

In addition, we looked at two non-gaze-related behavioral measures: *total time on task* (measured from when the first set of search results appeared on the screen until participants announced they had finished), and *click accuracy* (whether a participant clicked on the "best" result in the first set of results).

General gaze characteristics for search results

Before describing the results of the various manipulations, we present some aggregate characteristics for how people view Web search results across all our search tasks and conditions. First, confirming previous findings [12], we found that people viewed search results in a roughly linear order. Most gaze activity was directed at the first few items; items ranked lower got users' attention last and least (Figure 3).

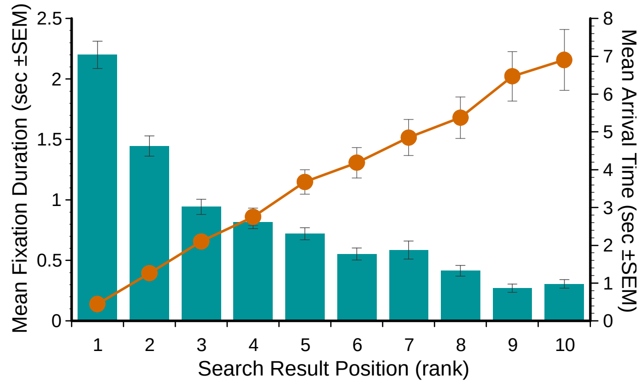

Figure 3. Mean fixation duration (bars) and mean time for gaze to arrive at each result (circles). As search results move downward in rank, it takes longer for searchers to arrive at them (upward trend of circles), and they spend less time looking at lower-ranked results (decreasing trend of bars). This figure sums across all search-results pages visited by participants. All error bars are ± standard error of the mean.

We also were interested in the number of items viewed before and after a selected item because this relates to how completely users search a set of results. That is, if users clicked on a result, on average how many other items above and below that item did they look at? Figure 4 shows that no matter which result they eventually clicked on, our participants usually looked at the first 3 or 4 search results. When they clicked on the first or second result, they still looked at the first 4 results. When they clicked on lower-ranked results, they usually had looked at most of the items ranked higher. Finally, participants went through about 8 results on a page before changing their queries without clicking on anything (indicated by "Requery"). With the exception of position 1, these results are very similar to findings reported by Joachims, et al. [12]. In their study, participants rarely looked at more than 1 or 2 items after the one they had clicked on, even when they had clicked on the first item.

A common observation in Web search is a "hub and spoke" pattern of exploration in which users go back and forth between a search results page and different target sites using the "back" button. We found that the distribution of

Figure 4. Mean number of search results looked at before users clicked on a result (above and below that result). E.g., if users clicked on result 5, on average they looked at almost all items above it and about 1.4 results below it.

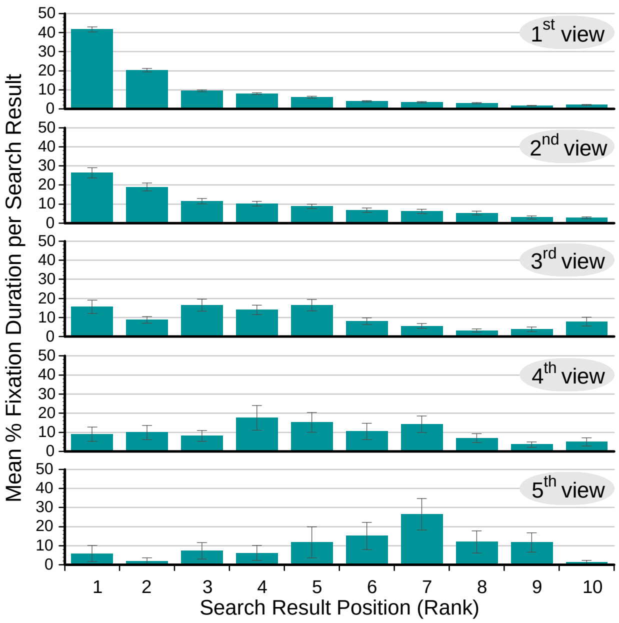

Figure 5. Mean percent of gaze fixation on each search result position broken down by visits to a given page of search results. Note that with subsequent viewing, searchers spend more time looking at lower-ranked search results.

users' fixations also changed with subsequent visits to the search page (see Figure 5). In the first visit, higher-ranked items got the most attention, as described in Figure 3. When a user returned to a results page for a second examination, higher-ranked items still received the most attention, but the slope of the fixation distribution decreased with proportionately more time spent on lower-ranked items. As the user returned to the page again, results 3, 4 and 5 became the main focus, and the focus steadily moved down the page with subsequent viewings.

Task Type & Snippet Length

To investigate the effect of task type and snippet length on how people search, we analyzed the following measures:

- Total time on task
- Total number of search results fixated for the task[2]
- Total summed duration of fixations on titles
- Total summed duration of fixations on snippets
- Total summed duration of fixations on URLs

For these measures, we performed a 2 (*Task Type*) x 3 (*Snippet Length*) x 2 (*Repetition*) repeated measures multivariate analysis of variance (RM MANOVA).

For *Task Type*, we found a significant main effect only for total time on task, $F(1,17)=54.7$, $p<0.001$. As in prior work [17], informational tasks took longer to finish than navigational tasks (78.1 s vs. 36.9 s).

[2] When we calculate the total number of results fixated for a task, we sum all the search results a participant looked at across all the result pages visited for that task. For example, if a participant looked at 5 results on the initial page, revised his query, and then looked at 6 on the new page, the total number of search results fixated for that task would be 11.

Table 3: F-values for *Task Type* x *Snippet Length*

Measure	F(2,34)	Sig. (p)
Time on task	4.4	0.02
# results fixated	5.6	0.01
Fix duration on titles	5.2	0.01
Fix duration on snippets	3.2	0.05
Fix duration on URLs	5.2	0.01

As expected, there was no main effect for *Repetition*, suggesting that search strategies did not change as a result of time and experience with the tasks. Nor was there a main effect on any measure for *Snippet Length*. This indicates that, averaging across both task types, changing the length of the query-dependent snippet had no effect on any measure of participants' search behavior.

However, when we looked at the interaction between *Task Type* x *Snippet Length*, we found significant effects for all 5 measures analyzed (see Table 3).

Figure 6 illustrates the mean time on task for each task type as we varied the snippet length. For navigational tasks, the time on task remained the same for short- and medium-length snippets but increased by 10 seconds for long snippets. In contrast, informational tasks showed an *improvement* in task time of 24 seconds with long snippets.

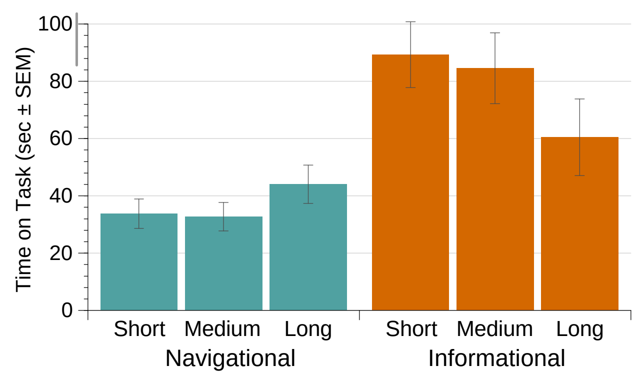

Figure 6. Mean time to complete search task for each task type, broken down by snippet length.

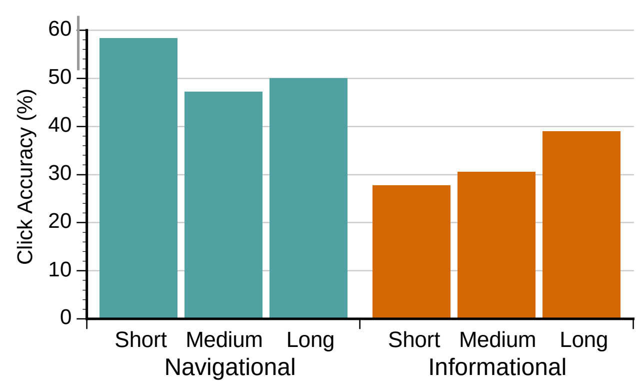

Figure 7. Accuracy in clicking the "best" result for each task type, broken down by snippet length.

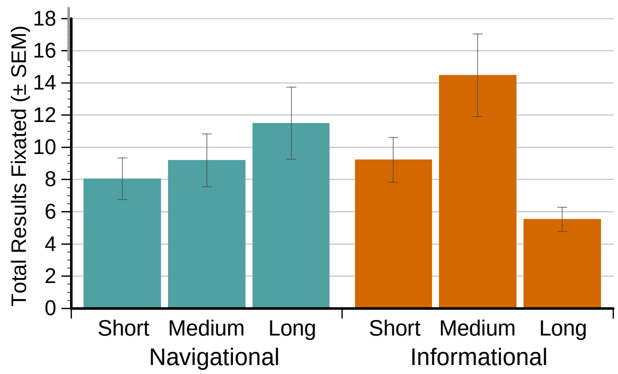

Figure 8. Mean number of search results fixated for each task type, broken down by snippet length.

Table 4. Mean fixation duration (with SEM) for each componenent of search results, broken down by task type and snippet length.

	Navigational		
	Short	Medium	Long
Title	3.36 (0.44)	4.31 (0.95)	5.49 (1.27)
Snippet	2.72 (0.52)	5.10 (1.29)	7.85 (1.90)
URL	2.93 (0.54)	3.25 (0.77)	3.32 (0.71)
	Informational		
	Short	Medium	Long
Title	5.56 (1.09)	6.68 (1.68)	3.61 (0.81)
Snippet	4.38 (0.82)	7.51 (1.54)	6.54 (1.34)
URL	3.79 (0.76)	4.16 (0.9)	1.47 (0.39)

If we focus only on informational tasks, this result supports the notion that more information in the snippet may help searchers determine whether a given site is likely to have the information they are interested in. To investigate this further, we also looked at the accuracy of our searchers' selections on the first query page where we know what the "best" result is. Figure 7 shows that as snippet length increased the accuracy of clicks for informational queries increased from 28% to 39%. Because of the small number of observations, statistical tests for this difference were not significant, but we believe that the trends provide converging evidence further supporting our hypothesis.

In stark contrast, Figures 6 and 7 indicate that increasing the snippet length had exactly the opposite effect for navigational tasks. Long snippets increased the total time on task, and snippets of even medium length were associated with a *drop* in accuracy from 58% to 47%. In sum, users performed best on navigational tasks with short snippets and best on informational tasks with long snippets.

To better understand what might be driving this overall effect on task performance, we looked at the gaze measures provided by eye tracking. Figure 8 shows that when searchers were given short snippets, they looked at about the same average number of search results independent of task type (about 8 or 9 search results). However, when they were given long snippets, searchers looked at more results when doing navigational tasks (about 12 search results). This was exactly the opposite for informational tasks, where searchers looked at one third *fewer* results with long snippets than when they were given short snippets.

This finding presents a puzzle. It seems plausible that increasing the amount of information on the search-results page would result in users looking at fewer results simply because there is more information to read; more lines of text in each result means that fewer results will be visible without scrolling. But, why would adding more information cause a user to look at *more* results? And why would this effect be task-dependent? One possible explanation is that adding more information would lead searchers to be more thorough, spending more time with search results because the information density is higher. They simply spend more time reading the results and less time reading Web documents. But if this were true, we should find that users look at more results in both informational and navigational tasks. What's going on here?

One possibility is that, because the goal of navigational tasks is locating a specific site, the information provided by contextual snippets is much less relevant for navigational than for informational tasks where details related to site content, authority, etc., are more important. In contrast, URLs may be proportionately *more* relevant for navigation because they are directly related to the location of target sites. If this were true, we would expect that searchers would spend proportionately more time looking at the URL in navigational than informational tasks. This was true in our study, but the difference was small: across all snippet lengths, people spent 25% of their time looking at the URL in navigational tasks vs. 22% in informational tasks. However, if we break this down by snippet length, a pattern begins to emerge. Figure 9 shows the relative proportion of total fixation duration for each search result component (title, snippet and URL) broken down by snippet length and task type (for reference, the mean fixation duration for each condition is shown in Table 4).

Figure 9 shows that as we increased the snippet length, the relative time spent looking at the snippet increased for both task types. However, while the proportion of time looking at the title stayed roughly constant, the increase in time

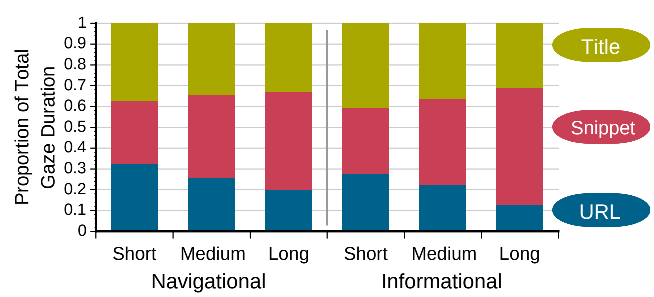

Figure 9. Proportion of total fixation duration for each component of search results, broken down by task type and snippet length. As snippet length increases, the relative proportion of gaze devoted to the URL decreases.

looking at the snippet came primarily at the cost of looking at the URL. This decrease was particularly dramatic for the informational tasks, but it was true for navigational tasks as well. Figure 9 suggests that when our participants looked at search results with long snippets, they read them, whether or not the snippets were relevant to their task.

Post-experimental questionnaire

After the experiment, participants answered a short questionnaire with questions about demographic information and their Web search experience. In general, our participants appeared to be quite savvy at Web search; most reported that they typically search the Web at least once a day, and all were familiar with and had used a variety of different search engines. Google was by far the most popular search engine, but several participants also reported using Yahoo! as their primary search engine. One reported using the AOL default engine as his primary search engine.

As part of the questionnaire, participants were given several 7-point Likert-scale questions of the nature, “Click 1 if you completely disagree, 7 if you completely agree and 4 if you neither agree nor disagree.” Of particular interest were answers to the following questions:

“The search terms automatically selected for each task were usually close to what I would have entered myself for that task.” For this question, the median score was 6 and the mean was 5.8. This was important, because we were concerned that the query terms we created would be so different from what users would generate on their own that their behavior would be unnatural. These scores suggest that for most of our participants, the query terms were reasonably close to what they would spontaneously generate.

The next two questions were also very interesting: *“When I'm searching the Web, I often look at the URL of each search result to help me decide if the page will be useful.”* And: *“When I'm searching the Web, I usually read the snippet (text under the title) to help me decide if the page will be useful.”* For these questions, the median scores were 7 and 6 respectively, and the means were 6.4 and 6.2. These answers suggest that our participants deliberately use various elements in the search results to help them find what they are looking for. We were particularly surprised to see the overwhelming endorsement of the URL because this is often characterized as a “power-user” feature that is used by only a small percentage of users.

DISCUSSION AND DESIGN IMPLICATIONS

The experiment described above presents designers of Web search engines with something of a dilemma. Our results showed that changing the user interface of Web search results by varying the length of the query-dependent contextual snippet had opposite effects on task performance depending on what a user was trying to do. For navigational tasks, optimal performance occurred with short snippet lengths, while for informational tasks, long snippets helped the most.

The informational needs from the search results were very different for the different task types in our study. All of our navigational tasks required participants to find a specific destination Web site. For these queries, the URL was likely to be a very useful source of information. This is not to say that the title and snippet were irrelevant—indeed in the condition that proved best for these tasks, searchers looked at the title, snippet and URL almost equally (Figure 9). In contrast, the URLs of search results were probably much less relevant for our informational tasks, because these tasks could be answered by any of a number of Web sites. So long as the destination site referenced by the search result looked authoritative and contained sufficient information scent, the searcher could be satisfied going to the result and looking for an answer from there.

If we take gaze fixation as a proxy for users’ attention, we can start to explain what is happening. For both task types, as the snippet length gets longer, attention to the snippet also increases. However, this increase comes at the cost of relative attention to the URL. The proportion of attention on the title also decreases somewhat, but the decrease is quite small. For informational tasks, where the URL is less relevant, the cost in task performance and information scent is minimal; the attention paid to the longer snippets more than makes up for any cost from missing the URL. In navigational tasks, the information in the longer snippets is not as relevant, and the cost of the lack of attention to the URL is more acute. Even though the total amount of time spent looking at URLs did not vary much for navigational tasks (see Table 4), we believe that the increased amount of attention to the longer snippets interfered with the information located in the URL, decreasing the URLs’ relevance for our users.

If users applied attention equally to each informational element in the search result, more information in the snippet could actually decrease their certainty that a result was the target, and they would continue looking at other results. That is, as more information is included in the results, users may unconsciously down-weight the relevance of URLs for their decisions. When multiple results have rich snippets, it would be more difficult to decide which result is the target in navigational tasks, but not in informational tasks, where the goal is any Web site likely to have the answer. This suggests that users may not consciously realize the benefit they receive from URLs and do not strategically devote attention to different parts of results depending on their task; they simply use what they are given.

This hypothesis is testable in a few ways. First, we could perform a similar study, this time removing URLs from the results. This would address the hypothesized importance of the URL for different task types. A more subtle variation would be to carefully choose a subset of navigational targets comprising Web sites that have URLs with little information (e.g., hosted by a generic provider, or with very long GUIDs in them). In this case, we would expect increased snippet lengths to have little effect (or perhaps a *positive* effect) on performance because URLs simply aren't useful for these tasks.

Another way to explore this hypothesis is to vary the attentional salience of different elements in the results list. If the above explanation is correct, we should be able to improve performance in navigational tasks by increasing the attention to navigational information such as URLs. Conversely, we could harm performance by emphasizing other, pseudo-relevant information. In either case, performance should directly correlate with the amount of gaze devoted to (ir)relevant information.

Our results suggest that for a substantial fraction of queries in Web search (informational tasks), extended snippets are useful. Despite users' having to scroll more, accuracy and task times were improved, and users actually looked at fewer total items. For another large class of Web search tasks (navigational), long snippets are problematic. However, our results suggest several possible solutions to this problem. In search results provided by MSN Search (and all major search engines), the URL is always placed at the bottom of each search result, immediately following the snippet. As seen in Figure 9, when the snippet is only a single line, all three elements receive almost equal attention as searchers linearly scan the results from top to bottom. But as the snippet length grows, searchers begin to lose the URLs in the mass of text. It would be very interesting to place the URL below the title, immediately above the snippet. This would guarantee that as users scan the results they would always see the URL before the snippet. When the URL is an important navigational aid, it would easily be seen; likewise, a single line of stylized text would be fairly easy to ignore if the useful information is in the snippet below it.

Another solution would be to radically alter the design of the results presentation to interrupt the linear top-to-bottom scanning. One might divide the display, placing the snippet in a dedicated pane to the right of the title, URL and other metadata. This would de-emphasize the snippet, but it would still be available for detailed examination.

Both of the above solutions assume an "all things for everyone" design in which search engine providers present results in a single style for all queries. Another solution would be to provide different information for different kinds of queries. This might involve an explicit gesture of intention from the user (e.g., in the trivial case, a button) or using automatic classification [14, 15]. If the search provider is able to determine reliably that a user is engaged in an informational task, it could provide results with richer content. Likewise, for clearly navigational tasks, it could minimize such content while emphasizing navigational information.

Our results also might have implications for search domains outside of Web search. There are many domains in which informational search is the primary activity (e.g., medical and academic databases). For these domains, our results suggest that long contextual snippets can greatly improve a user's search experience. Similarly, there are other domains, such as directory searches, where navigational tasks clearly dominate. For these domains, users may be better served by brief snippets and search results that emphasize navigational information such as the URL or other location context.

CONCLUSIONS AND NEXT STEPS

We presented a study using eye tracking techniques to investigate user strategies for Web search. In particular, we looked at how varying the amount of information in Web search results affected user performance on two kinds of search tasks. We found that as we increased the length of the query-dependent contextual snippet in search results, performance improved for informational queries but degraded for navigational queries. Our eye-tracking results suggest this difference in performance was due to the fact that as the snippet length increased, users paid more attention to the snippet and less attention to the URL located at the bottom of the search result.

Web search is a very attractive domain for the use of eye tracking techniques, and we believe this study is only a prelude to a wide range of studies in UI for information retrieval. For example, the experiments outlined in the previous section would provide excellent information about how users deploy their attention when they view search results. Similarly, it would be interesting to verify whether or not moving the URL above the snippet would improve users' experience in navigational search. There are many kinds of metadata that are potentially useful for Web search. How can this information be presented to users in a way that is complementary to existing information in search results?

This study raises theoretical questions about how our results might be situated with respect to information foraging theory [10]. It would be an interesting exercise to fit our data to the concept of information scent.

The future of Web search interfaces probably will be very different from what we see today [20]. Studies such as those we have outlined here can help to inform what those interfaces will look like. In addition, we would like to perform similar studies in other search domains to see whether our findings apply outside of Web search (e.g., search in corporate intranets, medical databases, personal desktop indices, etc.).

ACKNOWLEDGMENTS

We thank Susan Dumais, Dan Liebling and Muru Subramani for their extensive assistance and intellectual horsepower. In addition we would like to thank all the participants who spent an hour searching the internet on our made-up tasks.

REFERENCES

1. Broder, A. A taxonomy of web search. *SIGIR Forum, 36*, 2(2002), 3-10.
2. Brumby, D.P. and Howes, A. Good enough but I'll just check: Web-page search as attentional refocusing. *Proc. 6th Int'l Conference on Cognitive Modeling*, Lawrence Erlbaum (2004), 46–50.
3. Card, S.K., Pirolli, P., Van Der Wege, M., Morrision, J.B., Reeder, R.W., Schraedley, P.K. and Boshart, J. Information scent as a driver of web behavior graphs: Results of a protocol analysis method for web usability. In *Proc. CHI 2001*, ACM Press (2001), 498–505.
4. Chi, E., Pirolli, P., Chen, K., & Pitkow, J. Using information scent to model user information needs and actions and the Web. In *Proc. CHI 2001*, ACM Press (2001), 490–497.
5. Cutrell, E., Dumais, S.T., & Teevan, J. Searching to eliminate personal information management. *In Communications of the ACM (Special Issue: Personal information management), 49* 1(2006), 58-64.
6. Cutrell, E., & Guan, Z. Eye tracking in MSN Search: Investigating snippet length, target position and task types. *Microsoft Technical Report, MSR-TR-2007-01*. ftp://ftp.research.microsoft.com/pub/tr/TR-2007-01.pdf. (2007).
7. Dumais, S.T., Cutrell, E. and Chen, H. Optimizing search by showing results in context. In *Proc. CHI 2001*, ACM Press (2001), 277-284.
8. Guan, Z., and Cutrell, E. An eye tracking study of the effect of target rank on Web search. In *Proc. CHI 2007*, ACM Press (2007).
9. Hornof, A.J., and Halverson, T. Cognitive strategies and eye movements for searching hierarchical computer displays. In *Proc. CHI 2003*, ACM Press (2003), 249-256.
10. Hotchkiss, G., Alston, S. and Edwards, G. Eye Tracking Study. http://www.enquiro.com/eyetrackingreport.asp.
11. Jansen, B.J. and Pooch, U. A Review of Web Searching Studies and a Framework for Future Research. *Journal of the American Society of Information Science and Technology, 52* (2000), 235–246.
12. Joachims, T., Granka, L., Pan, B., Hembrooke, H., and Gay, G. Accurately interpreting clickthrough data as implicit feedback. In *Proc. SIGIR 2005*, ACMPress (2005), 154-161.
13. Klöckner, K., Wirschum, N. and Jameson, A. Depth- and breadth-first processing of search result lists. In *Ext. Abstracts CHI 2004*, ACM Press (2004), 1539-1539.
14. Lau, T. and Horvitz, E. Patterns of search: analyzing and modeling web query refinement. *Proc. Seventh Int'l Conference on User Modeling*, Springer-Verlag (1999), 119–128.
15. Lee, U., Liu, Z., and Cho, Junghoo. Automatic identification of user goals in web search. In *Proc. WWW 2005*, (2005), 391-400.
16. Lin, J., Quan, D., Sinha, V., Bakshi, K., Huynh, D., Katz, B., and Karger, D.R. The role of context in question answering systems. In *Ext. Abstracts CHI 2003*, ACM Press (2003), 1006-1007.
17. Lorigo, L., Pan, B., Hembrooke, H., Joachims, T., Granka, L., and Gay, G. The influence of task and gender on search and evaluation behavior using Google. *Info. Processing and Management: an Int'l Journal*. *42*, 4 (2006), 1123-1131.
18. Mat-Hassan, M. and Levene, M. Associating search and navigation behavior through log analysis. *J American Society for Information Science and Technology, 56* (2005), 913-934.
19. Pratt, W. and Fagan, L. The usefulness of dynamically categorizing search results. *J American Medical Informatics Association*, *7* 6(2000), 605-617.
20. Rele, R.S. and Duchowski, A.T. Using eye tracking to evaluate alternate search results interfaces. In *Proc. HFES,49th Annual Meeting* (2005).
21. Rose, D.E. Reconciling information-seeking behavior with search user interfaces for the web. *J American Society for Information Science and Technology, 57* (2006), 797-799.
22. Rose, D.E. and Levinson, D. Understanding user goals in Web search. In *Proc. WWW 2004*, (2004), 13-19.
23. Silverstein, C., Henzinger, M., Marais, H. and Moricz, M. Analysis of a very large AltaVista query log. *SRC Technical note #1998-14*. http://gatekeeper.dec.com/pub/DEC/SRC/technical-notes/abstracts/src-tn-1998-014.html. (1998).
24. 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. In *Proc. CHI 2004*, ACM Press (2004), 415-422.
25. Yee, K.P., Swearingen, K., Li, K., and Hearst, M. Faceted metadata for image search and browsing. *Proc. SIGCHI 2003*, ACM Press (2003), 401-408.
26. Zhang, J. and Marchionini, G. Evaluation and evolution of a browse and search interface: relation browser++. In *Proc. CHI 2005*, ACM Press (2005), 179-188.