An Eye Tracking Study of the Effect of Target Rank on Web Search

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

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

ABSTRACT
Web search engines present search results in a rank ordered list. This works when what a user wants is near the top, but sometimes the information that the user really wants is located at the bottom of the page. This study examined how users' search behaviors vary when target results were displayed at various positions for informational and navigational tasks. We found that when targets were placed relatively low in the first page of search results, people spent more time searching and were less successful in finding the target, especially for informational tasks. Further analysis of eye movements showed that the decrease in search performance was partially due to the fact that users rarely looked at lower ranking results. The large decrease in performance for informational search is probably because users have high confidence in the search engine's ranking; in contrast to navigational tasks, where the target is more obvious from information presented in the results, in informational tasks, users try out the top ranked results even if these results are perceived as less relevant for the task.

Author Keywords
Web search, eye tracking, target position, trust.

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

INTRODUCTION
With the increase in volume of digital information, search has become one of the most efficient ways to find what users are looking for. Various search engines or search services have been launched to help users find information stored on World Wide Web, inside corporate networks, or on personal computers.

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

When a user searches for information with a search engine, its effectiveness is determined by whether it gives back relevant results. Most search engines display results in a rank-ordered list, with the highest ranked result placed on top and others ordered below that.

Although this is efficient when the first few results displayed in the list are the most relevant, such ranking can be problematic when these results are not what users are looking for. Past studies [6] have shown that people often choose the first few results on the top of the list and ignore the rest. It was observed that users often click on the first item even if the second is more relevant. In addition, users may simply change their queries when the first few results are not promising, even though some results further down the list might well satisfy their search goals. This leads us to ask: how does the ranking (as determined by a search engine) of the results affect how people search? Do they blindly follow the search engine's ranking or make their own judgment of results based on information they see? What happens when the user's goal is not included at the top of the search result list?

EXPERIMENT
To investigate how people search for information when the best result is not on top, we designed a study that varied the absolute rank position of the "best" search result for each task. We used eye tracking to record what people looked at during search. Eye tracking technologies have been widely used as a proxy for users' attention. Eye movement data helps us understand where people invest attention, and in what order before they make a selection[5].

Design
The design of the experiment crossed *Task Type (2)* x *Target Position (6)* as two within subject factors. Two types of search tasks (navigational and informational tasks) identified in the literature [1] were used in this study. In *navigational* tasks, users were asked to find a specific website or homepage for the task; the goal was simply to get to their destination. In *informational* tasks the goal was to acquire some kind of information irrespective of where it was located. The target result was displayed at six positions (1, 2, 4, 5, 7, and 8) for each task. The study also systematically varied the length of the descriptive text. For the results related to the snippet length, please see [3].

Each participant completed a sequence of 12 search tasks (2 task types x 6 target positions). Tasks were randomly ordered. Each of 12 search tasks (6 of each task type) was counterbalanced across participants such that every task was seen with every target position. Each search task consisted of a brief motivation statement, task description, and a hyperlink containing a predefined query that would launch a search when clicked. The queries were designed such that the task could be completed with a site presented in the 10 results returned. All the results were directly retrieved from the search engine (MSN search) using the associated query in August 2006.

Apparatus

All Web search results were received from a special server for MSN Search (http://search.msn.com). The position of the target result was manipulated using a proxy. Eye tracking was performed using the Tobii x50 eye-tracker (50Hz) paired with a 17" LCD monitor (96 dpi) set at a resolution of 1024x768. An integrated log of eye movement data, user events and Web pages visited was collected.

Participants

Twenty-two participants ranging in age from 18 to 50 years old were recruited for this study. Of these, 4 participants were excluded due to stability problems with the eye tracking, leaving us with 18 participants (11 male). All participants were moderately experienced at Web search, and all were familiar with several different search engines. None of them had experience using an eye-tracker.

Procedure

At the beginning of the study, the eye-tracker was calibrated for each participant and they were given a practice query to get familiar 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 the participant clicked on the target page, confirmed it was the desired site and vocally announced that they had found the website or information requested. Following completion of all search tasks, participants answered a short questionnaire about their experiences in the study, and provided demographic information.

For a more detailed description of the experiment design and study setting, including the complete list of used search tasks (queries), a screenshot of a search results page, the setup of the proxy, and the generation of the areas of interest (AOIs), please see [3] [4].

RESULTS

In this paper, we focus our analysis on users' task performance and gaze fixation measures. The fixations were aggregated from gaze points with a minimum threshold of 100 ms in areas of interest. Each individual search result was considered an *area of interest*. Since the target position was manipulated on the first page in the experiment where results were pre-cached, all the measures are on participants' first encounter of the first page, except the total time on task. Total time on task and fixation measures were analyzed using 2 (*Task Type*) x 6 (*Target Position*) repeated measures analysis of variance (ANOVA). The click accuracy was analyzed using a chi-square analysis.

General Effects on Task Performance

We found a significant main effect of *Target Position* on the total time on task ($F(5,85)=3.544$, $p=.006$). This indicates that people spent significantly more time on a task when the target was displayed at a lower position. We also found a main effect for *Task Type*, $F(1,17)=54.718$, $p<0.001$, confirming what [7] found. There was no significant interaction between *Target Position* and *Task Type*.

While participants took more time finishing tasks when the target position moved down, it didn't help them make accurate selections. A chi-square analysis on the number of accurate clicks showed a significant effect for target position ($\chi^2(5)=58.5$, $p<0.001$). The click accuracy rate dropped from 84% (average of 78% and 89%) to about 11% when the target was displayed at position 8 (Fig. 1).

Figure 1 shows that for navigational tasks, people had the highest click accuracy rate when the target was in the first 2

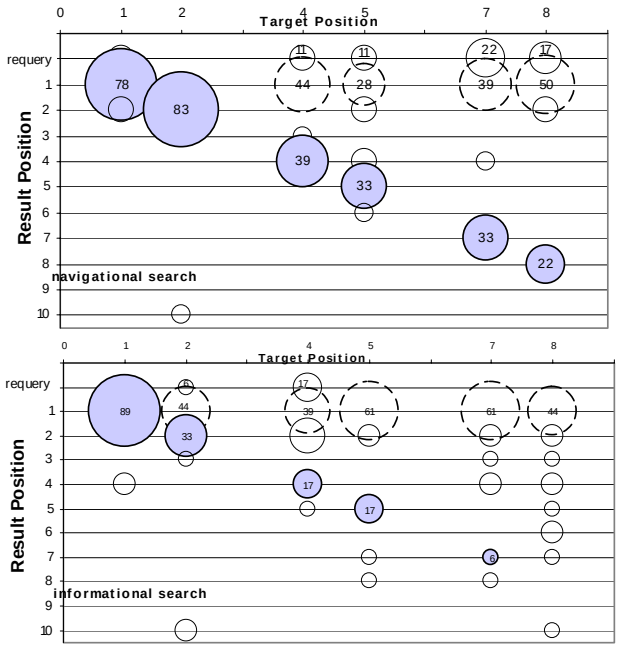

Figure 1: chance of clicking on search results break down with target position. The numbers inside the bubble indicate the chance (in %) that the result was clicked (e.g. when the target position is 2 for navigational search, 83% of participants clicked on result 2, which is the target result.) The shadowed bubbles indicate the target results. The bubble with a dashed border indicates the first result. Bigger bubble indicates a larger probability of clicking at the result at its particular position, which is also shown with a number inside the bubble.

positions (78%, 83%). With the target at position 4, 5, 7, and 8, click accuracy dropped to 39% or less. For informational tasks, the effect of target position on click accuracy was much more dramatic. Participants correctly selected the target less than 20% of the time when the target was below position 2, and not a single participant correctly selected the target when it was at position 8.

A closer look at where people clicked (Fig. 1) shows another interesting phenomenon. Across both task types, when targets were placed at lower positions (4,5,7, or 8), participants frequently clicked on the first result (average of 46% of the time). For navigational tasks, participants issued a new query without clicking any result 15% of the time. However, for informational tasks, they rarely re-queried without clicking on anything (4%). More than half the time they chose the first result, or else clicked around on other results.

It is not surprising to see that when the target position was moved to the lower part of the result list, participants spent more time on the tasks yet achieved poorer accuracy. However, we did not expect such a dramatic effect, particularly for informational search for which participants achieved less than 10% accuracy. We hypothesized there might be two reasons for the general decrease of click accuracy across different tasks and the dramatic effect on informational search:

1) Since participants rarely went through the whole result list, they never saw the target result when it was placed at a low position, especially for informational search. This could be tested by looking at the number of results people fixated upon.

2) Alternatively, participants may have seen the target result for both navigational search and informational search, but they did not feel the results at lower positions were as compelling as others. This could be tested by looking at the effect of task type and target position on fixation duration (an indicator of participants' attention.)

Examination of the gaze distribution may help us to understand the dramatic difference in selecting target results depending on their ranks.

Did Users Look at Target Results?

A 2 x 6 ANOVA (see above for model) on the number of results participants fixated upon within the first page shows that there was a main effect of *Target Position*, $F(5, 85)=4.958$, $p=.011$. Participants went through more results (for position 1, mean=3.47, SE=.409; for position 8, mean=6.06, SE=.572) in order to complete the task when the target was placed lower. This indicates that participants sensed the fact that the top results were not correct and felt difficulties in finding the target when it was placed lower. No significant effect was found for *Task Type* or for the *Task Type x Target Position* interaction.

We looked further at the accumulated times that people fixated upon the target results (Fig. 2). For navigational search, everyone looked at the target result when it was the first result (100%). When the target was position 2, this dropped to 89%, then to 72% for 4th, 56% for 5th, 7th and 8th. For informational search, the chance of looking at the target result dropped a little faster from over 90% for position 1 and 2, to 22% for position 8 (see Table 1).

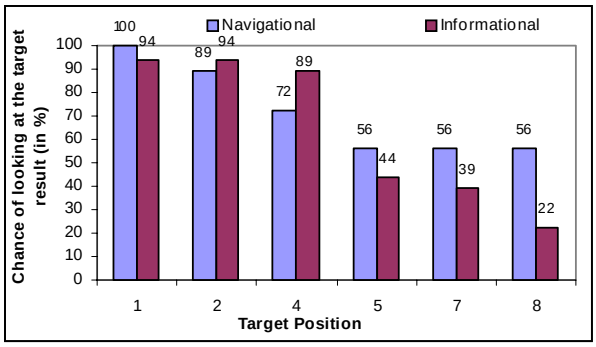

Figure 2: Chance of looking at target results (e.g. when the target position is 1, everyone look at the target result for navigational search (100%), and 89% of them look at the target for informational search.

Table 1: % of people looked at the target result (% of people clicked at them) for navigational and informational search.

Target position	1	2	4	5	7	8
Navigational	100(78)	89(83)	72(39)	56(33)	56(33)	56(22)
Informational	94(89)	94(33)	89(17)	44(17)	39(6)	22(0)

This result supports the first hypothesis above, that the decreased probability of clicking on the target is related to the probability of looking at the target: if a user doesn't see a result, he won't click on it. However, this still doesn't explain the dramatic decrease in click accuracy for informational search: participants were likely to look at the targets at positions 2 and 4, but were extremely reluctant to click on them (see Table 1). Is this because participants allocate less attention to lower results even though they looked at them (hypothesis 2)? Further analysis of fixation duration rejects this possibility.

How Much Attention Did Users Invest on the Target Results?

Our analysis of how long people looked at search results when the target results were at different positions leads us to suspect other reasons (e.g. high confidence in search engine) to explain people's reluctance to select the target results during informational search.

A repeated measures ANOVA found a main effect for *Target Position*, $F(5,85)=7.06$, $p<0.001$. The average fixation time on target results decreased with lowering target position. However, we found no main effect of *Task Type* and no *Task Type* x *Target Position* interaction. This means that people looked at the targets in the same way for navigational and informational tasks. Furthermore, the fixation duration on targets at lower positions decreased at the same rate for navigational and informational tasks.

This result shows that for informational tasks, people looked at the same number of lower-ranked results as they did for navigational tasks, but they clicked much less frequently on them. Figure 3 also indicates that people often lingered on the target results for informational search even though they didn't click on them(suggesting a sensitivity to information scent [2]). This suggests that users trust the search engine more for informational search or invest less scrutiny in judging the results with higher rankings. Eventually they are more likely to choose the top few results to try them out in spite of their lower objective relevance to the task. In the post-questionnaire, several responses from participants on their expectations for search results also speak to this effect: they highly agreed with the statement "I expect the information I'm looking for to be in the top five results" (mean=5.78, SE=.94, on a 7 point Likert-scale). Participants showed no preference on the statement "I often scroll to the bottom of the first page of search results looking for what I want" (mean=4.06, SE=1.63.)

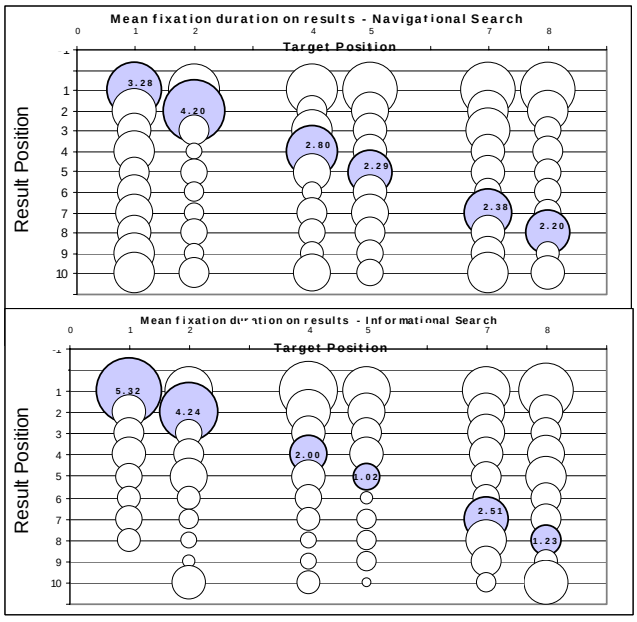

Figure 3: Fixation duration on results when people looked at them, break down by target position. The shadowed bubbles indicate the target results.

CONCLUSION

Web search engines have become commoditized tools for finding information in our daily lives. Most search engines display search results in a rank ordered list, with the closest matched results placed on top and others ordered below that. However, this display has a potential side effect: users may not utilize more relevant results that are displayed at lower positions on the list.

This study showed that people spend more time on tasks and are less successful in finding target results when targets were displayed at lower positions in the list. When people could not find the target results for navigational search, they either selected the first result, or switched to a new query. For informational search, people rarely issued a new query and were more likely to try out the top-ranked results despite their lower objective relevance to the task.

Further eye movement analysis suggests that the uniform decrease in click accuracy for both navigational and informational search may be due to the fact that people only go through the results on the top of the list. The analysis also showed that the large decreases in performance for informational search could be a result of 1) a decreased probability for looking at lower results, and 2) possible strong confidence in search engine relevance ranking even though people clearly see target results at lower positions. People are more likely to deprecate their own sense of objective relevance and obey the ranking determined by the search engine. This result implies that the search engine could show variety of different search results where "best" ranking is not clear so that users could have an accurate estimation of the relevance of results and then behave accordingly.

This study empirically studied how people's attention was distributed across search results when the target was systematically manipulated to be displayed at different positions. Further studies and analyses include examining how people distribute their attention across different parts of results (e.g. title, snippet, or URL), and making design adjustments on search interface to encourage people to explore more results if the top results are not compelling.

ACKNOWLEDGMENTS

We greatly appreciate Susan Dumais for her insightful comments and Dan Liebling and Muru Subramani for their technical help.

REFERENCES

1. Broder, A., A taxonomy of web search. *SIGIR Forum*, (2002). 36, 2, 3-10.

2. Card, S.K., Pirolli, P., Wege, M.V.D., Morrison, 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'01*. (2001). 498-505.

3. Cutrell, E. and Guan, Z. What are you looking for? An eye-tracking study of information usage in Web search. In *ACM CHI'07*. (2007). TBD.

4. Cutrell, E. and Guan, Z., Eye tracking in MSN Search: Investigating snippet length, target position and task types. *Microsoft Technical Report*, (2007)MSR-TR-2007-01.

5. Hornof, A.J. and Halverson, T. Cognitive strategies and eye movements for searching hierarchical computer displays. In *ACM CHI'03*. (2003). 249-256.

6. Joachims, T., Granka, L., Bing Pan, and Gay, G. Accurately Interpreting Clickthrough Data as Implicit Feedback. In *ACM SIGIR'05*. (2005). 154 - 161.

7. 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*, (2006). 42, 4, 1123-1131.

The columns on the last page should be of approximately equal length.