Eye-Tracking Analysis of User Behavior in WWW Search

Laura A. Granka
Cornell University
Human-Computer Interaction Group
lag24@cornell.edu

Thorsten Joachims
Cornell University
Department of Computer Science
tj@cs.cornell.edu

Geri Gay
Cornell University
Human-Computer Interaction Group
gkg1@cornell.edu

ABSTRACT
We investigate how users interact with the results page of a WWW search engine using eye-tracking. The goal is to gain insight into how users browse the presented abstracts and how they select links for further exploration. Such understanding is valuable for improved interface design, as well as for more accurate interpretations of implicit feedback (e.g. clickthrough) for machine learning. The following presents initial results, focusing on the amount of time spent viewing the presented abstracts, the total number of abstract viewed, as well as measures of how thoroughly searchers evaluate their results set.

Categories and Subject Descriptors
H.5.2 [**User Interfaces**]: *Evaluation/methodology*, H.3.3 [**Information Search and Retrieval**]: *Search process,* H.3.5 [**Online Information Services**]: *Web-based services*

General Terms
Human Factors, Experimentation, Measurement

Keywords
Eye-Tracking, Implicit Feedback, WWW Search

1. INTRODUCTION
How do users interact with the list of ranked results of WWW search engines? Do they read the abstracts sequentially from top to bottom, or do they skip links? How many of the results do users evaluate before clicking on a link or reformulating the search? The answers to these questions will be beneficial in at least three ways. First, they provide the basis for improved interfaces. Second, they suggest more targeted metrics for evaluating the retrieval performance in WWW search. And third, they help interpreting implicit feedback like clickthrough and reading times for machine learning of improved retrieval functions [2]. In particular, better understanding of user behavior will allow us to draw more accurate inferences about how implicit feedback relates to relative relevance judgments.

The following presents the results of an eye-tracking study that we conducted. Previous studies have analyzed directly observable data like query word frequency [6]. However, unlike eye-tracking, these measurements can at best give indirect evidence of how users perceive and respond to the search results.

SIGIR'04, July 25–29, 2004, Sheffield, South Yorkshire, UK.
ACM 1-58113-881-4/04/0007.

To the best of our knowledge, only one previous study has used eye-tracking in the context of information retrieval evaluation [5]. This study attempted to use eye movements to infer the relevancy of documents in the retrieval phase of an information search. The researchers linked relevancy judgments to increases in pupil diameter, as a larger diameter typically signifies high interest in the content matter. However, the sample size and search tasks in this experiment were not robust enough to generate predictable patterns of user search and scanning behavior, which is what our study is able to attain.

2. EYE-TRACKING
The research presented here seeks to obtain a more comprehensive understanding of *what* the searcher is doing and reading before actually selecting an online document. Ocular indices enable us to determine what abstracts a user is indeed viewing and reading, for how long, and in what order. Throughout the history of eye tracking research, several key variables have emerged as significant indicators of ocular behaviors, including fixations, saccades, pupil dilation, and scan paths [3]. Eye fixations are defined as a spatially stable gaze lasting for approximately 200-300 milliseconds, during which visual attention is directed to a specific area of the visual display. Fixations represent the instances in which information acquisition and processing is able to occur, and thus, fixations were the indices most relevant to this current evaluation [3]. Pupil dilation is typically used as a measure to gauge an individual's interest or arousal in the content they are viewing.

3. EXPERIMENT
Participants were undergraduate students of various majors at a large university in the Northeast USA. In total, 36 participants were recruited. Due to the inability of some subjects to be precisely calibrated, complete eye movement data was recorded for 26 of the subjects. The mean age of users was 20.3, with 19 males and 15 females. Nearly all subjects reported a high familiarity with the Google interface, with 31 users indicating that Google is their primary search engine.

Each participant was given the same ten questions to answer. Five of the questions are homepage-searches, the other five are informational searches [1]. The questions vary in difficulty and topic, covering travel, transportation, science, movies, local, politics, television, college, and trivia. Subjects were instructed to search as they normally would, and were not informed that we were specifically interested in their behavior on the results page of Google.

Data was recorded using an ASL 504 commercial eye-tracker (Applied Science Technologies, Bedford, MA) which utilizes a CCD camera that reconstructs a subject's eye position through the

Pupil-Center and Corneal-Reflection method. A software application accompanying the system was used for the simultaneous acquisition of the subject's eye movements. To perform analyses, "LookZones" were constructed around each of the ten results (title, abstract, and metadata) displayed on a Google results page.

4. RESULTS AND DISCUSSION

In all, our data consists of 397 queries. In the following we analyze all behavior before a user clicks on the first link, or exits the page otherwise. Further clicks are not considered in this paper. On average, it took participants 7.78 seconds to select a document (SE = .37). However, the time varies significantly between the 10 search tasks, from 5-6 seconds up to 11 seconds for the most difficult questions.

4.1 How does rank influence the amount of attention a link receives?

One of the valuable aspects of eye-tracking is that we can determine how the displayed results are actually viewed. Figure 1 shows the mean time users fixate on a presented abstract at that rank, as well as the number of clicks. Interestingly, the time is almost equal for links ranked 1 and 2. This is in contrast to the fact that users substantially more often click on the link ranked first. After the second link, fixation time drops off sharply. There is an interesting dip around result 6/7, both in the viewing time as well as in the number of clicks. Unlike for ranks 2 to 5, the abstracts ranked 6 to 10 receive approximately equal attention. This can be explained by the fact that typically only the first 5-6 links were visible without scrolling. Once the user has started scrolling, rank becomes less of an influence for attention. A sharp drop occurs after link 10, as ten results are displayed per page.

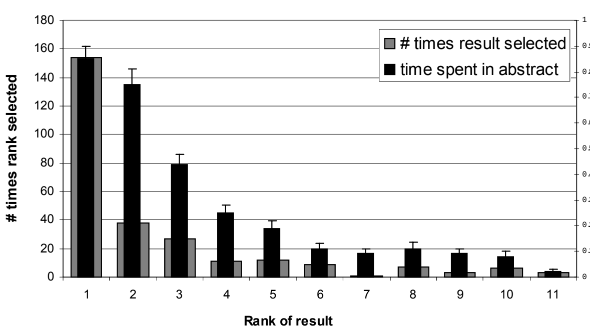

Figure 1. Time spent viewing each abstract with the frequency that abstracts are selected. Error bars are 1 SE

4.2 How do Users explore the List?

Particularly when observed user actions serve as implicit feedback about the performance of a retrieval system, it is important to know how thoroughly users evaluate the presented results before making a selection. For instance, if a user clicks on the third-ranked result, did she look at abstracts one and two? Did the user explore any links below? Figure 2 depicts how many results above and below the selected document users scan on average. Again, there is an interesting effect around before the page break. First, only one individual clicked on rank seven, which often fell directly below the page break. Secondly, users who selected the lower ranked documents viewed proportionately more abstracts overall. Finally, the number of links viewed below a click is low beyond rank 1, indicating that users do tend to scan the list from top to bottom.

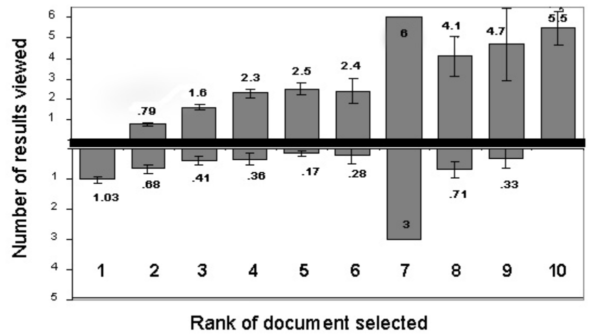

Figure 2. Number of abstracts viewed above and below the selected document. Error bars are 1 SE.

5. FUTURE WORK

In addition to further evaluations of the eye tracking data itself (e.g. for differences between question and users, as well as additional measures like pupil dilation), we are currently gathering relevance judgments for all abstracts and documents presented to the users. This will allow us to assess user behavior in relation to the relevance of document. For example, how accurately can users judge relevance of a document given the abstract in relation to fixation times? Do users tend to click on the most relevant link among the ones they observed? We will present those findings at the poster.

6. REFERENCES

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

[2] Joachims, T. Optimizing search engines using clickthrough data. Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD), ACM, 2002, pp 132-142.

[3] Rayner, K. Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124: 372-422, 1998.

[4] Salogarvi, J., Kojo, I., Jaana, S., and Kaski, S. Can relevance be inferred from eye movements in information retrieval? In Proceedings of the Workshop on Self-Organizing Maps (WSOM'03), Hibikino, Kitakyushu, Japan, September 2003. pp. 261-266.

[5] Silverstein, C., Henzinger, M., Marais, J., Moricz, M. Analysis of a very large AltaVista query log. Technical Report, Hewlett Packard Laboratories, Number SRC-TN 1998-014, Oct. 19, 1998.