ABSTRACT
Understanding the behaviors people employ when using a search system is important for developing systems that better serve users. More detailed knowledge of where searchers focus their visual attention is important to understanding how searchers use search results lists when reformulating queries. While observation of visual scanning has been used in the study of relevance judgments, to our knowledge the study reported here is the first to investigate visual attention to query terms. The study presented in this paper utilized eye-tracking to investigate the frequency with which people looked at (fixated on) their query terms when the terms appeared on search engine results pages. Results show that people fixated on some of the displayed query terms; however, they fixated on other words and parts of the page more frequently.

Keywords
Web search, eye-tracking, user studies

INTRODUCTION
Though there is much research on the relationship between visual scanning and the relevance of search results, less is known about how visual scanning is related to other stages of search such as query reformulation. Understanding the relationship between visual scanning and search stages allows for search system optimization. The study presented in this paper examined whether and how frequently searchers looked at their query terms when scanning a search engine results page (SERP) during Web search.

RELATED WORK
Prior research suggests that search behavior is affected by the presence of query terms on SERPs. Clarke, Agichtein, Dumais, and White (2007) found evidence that an item is more likely to be clicked when query terms appear in the item’s summary (the item’s representation on the results list), and that the greater the number of query terms the greater the probability. They found the effect to be largest when the terms appeared in the title of the summary. Further, the results of Yue, Patel, and Roehrig (2010) suggest that the query terms draw attention when they are bolded in the title. They found that, independent of the relevance of the item, searchers were more likely to click on a summary when it contained bolded query terms. Taken together, the above results support the notion that searchers look for their query terms on the SERP and that the presence of those terms is important to the searcher in determining which results will take them to useful sources. In a more recent study, Eickhoff, Teevan, White, and Dumais (2014) examined the words used in queries and found that for reformulations, twenty-five percent of terms added to queries appeared on a previously viewed SERP. The finding suggests that when a SERP does not contain sufficient relevant items, searchers’ visual attention is drawn to words other than their query terms. None of the studies discussed above used eye-tracking.

Eye-tracking studies have revealed patterns and bias in searchers’ visual attention when scanning a SERP before making a choice of where to click (Granka, Joachims, & Gay, 2004). While little is known about how these patterns relate to query reformulation, it has been proposed that searchers’ visual attention might be used as implicit feedback on relevance (Salojärvi, Kojo, Simola, & Kaski, 2003). Researchers have also proposed that eye-tracking be used for automated query modification. Hardoon, Shawe-Taylor, Ajanki, Puolamäki, and Kaski, (2007) suggested that the use of eye-tracking would enable gaze-based implicit reformulation where terms fixated upon after an initial query could be extracted and used in a new query. Progress on such ideas requires better understanding of the processes people employ while scanning SERPs and reformulating queries.

The study presented in this paper used eye-tracking to describe the visual scanning behavior of people using a Web search engine. The objective was to better understand how users interact with query terms they use.
assigned searches. In this section of the paper we describe the design, data collection methods, participants and procedures used.

**Design**

In order to control for variation due to search topics, search tasks were assigned to participants. Because Toms et al. (2008) found that their participants obtained more than half of their query terms from words provided in search task descriptions, our tasks were designed uniquely to ensure participants generated their own query terms rather than copying them from the task. For each task, the participant was given two keywords aligned with the specific intention of evoking an association between the terms. Both words were always nouns, and the first word could always perform a specific action on the second word. Participants were directed to find a website that provided information about the association, which was not given to them explicitly. Figure 1 shows the practice task given to participants; it aimed to evoke the concept of herding. Table 1 shows the noun pairs used in the eleven assigned search tasks and one practice task used in the study. Participants were instructed not to use either keyword in their queries. To control for order effects, task presentation was controlled automatically by Tobii Studio using a Latin square design. Task assignments were displayed on the eye-tracker screen and on printed paper.

**Data Collection**

Data were collected using a Windows 7 PC hosting Tobii Studio version 2.1.13 which was connected via a TCP/IP connection to the Tobii T60 eye-tracker used to collect gaze data. All gaze data and the participants' view of the eye-tracker screen (screengrabs) were captured, although only data corresponding to SERPs were used for analysis. All default Tobii Studio settings were used. The researcher cleared the browsing history between each participant session. After gaze data were collected, frequency counts and other derived data were produced following the steps described in the analysis section.

Figure 2 depicts the flow of items viewed during each session. Solid arrows indicate the direction of actions that always occurred. Dotted lines indicate optional actions. Rectangles denote items which were required to be viewed during the session and a circle indicates an item which was not required. The eye icon indicates the item for which recorded eye gaze data was later used in analysis.

**Participants**

Eighteen participants were recruited from an undergraduate research pool; all earned course credit for participation. All had either normal or corrected-to-normal vision and all were between the ages of 18 and 24. Based on education level and major areas of study, it can be inferred that no participants had any advanced training in searching. The exact level of search experience each participant had, however, was not known.

**Procedure**

Each session lasted up to forty-five minutes, though many participants took less time. All parts of the study were presented on the eye-tracker screen. A paper copy of each task was also placed next to the computer for reference during search. Each session was self-paced, with no time limit for any particular task. The researcher did not mention the purpose of the study.

After arriving at the session and receiving general instructions,

---

**Table 1. All noun pairs used in tasks.**

<table>
<thead>
<tr>
<th>Task Noun Pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>author and book</td>
</tr>
<tr>
<td>brush and cat</td>
</tr>
<tr>
<td>citizen and ballot</td>
</tr>
<tr>
<td>dog and sheep (practice)</td>
</tr>
<tr>
<td>eye and page</td>
</tr>
<tr>
<td>father and child</td>
</tr>
<tr>
<td>maid and house</td>
</tr>
<tr>
<td>match and wood</td>
</tr>
<tr>
<td>car and road</td>
</tr>
<tr>
<td>needle and thread</td>
</tr>
<tr>
<td>oven and food</td>
</tr>
<tr>
<td>tractor and seed</td>
</tr>
</tbody>
</table>

---

Figure 1. Image of the practice task, showing the form of all tasks and instructions given to participants.
detailed instructions were displayed, followed by a single practice task. For each task, participants saw the task description before submitting their first query and receiving a SERP. Participants then clicked to review Web pages found, with the option of submitting additional queries to search until a good page was found and selected as fulfilling the task. The process continued for each of the eleven assigned searches, after which the session ended. Most participants used multiple queries per task.

**ANALYSIS**

The Tobii system saved a screenshot of every Web page and SERP viewed during each session, along with data on the location and duration of each visual fixation on those pages. The analysis examined only fixations on SERPs. Data from the practice task SERPs and all non-SERP pages were excluded. For every SERP, the location of each query term was identified using the Tobii Area of interest (AOI) feature, which allows segmentation and identification of screen areas. If a fixation falls within an AOI, the system automatically labels the fixation. For our analysis, the first author drew AOI boundaries around each instance of every query term displayed on a SERP. Only major concept words and their variants were labeled; function words were excluded (see Table 2). For example, for the query “how to and voting and democracy” only the words voting and democracy and their root word variants (plural/singular noun, verb, adjective) were identified as AOIs. Query terms appearing in advertisements and query suggestions were included. AOIs were drawn on the entire SERP, including areas that received no fixations. The second author reviewed every SERP and AOI for identification errors and found only a one percent error rate; nearly all errors were query terms that were not identified, with very few non-query terms erroneously identified as AOIs.

Once all AOIs were drawn, the following values were computed for each participant and then again for each task.

- Query term fixations (QTF): the number of fixations on query terms
- Other fixations (OF): the number of fixations not on query terms
- Total fixations (TF): QTF + OF or the total number of all fixations
- Query fixation ratio (QFR): QTF/TF or the ratio of query term fixations to total fixations
- Total fixation ratio (TFR): OF/TF or the ratio of other fixations to total fixations
- Total query terms (TQT): the total number of query term as defined by AOIs
- AOI ratio (AOIR), QTF/TQT or the ratio of total query term fixations to the total number of query terms.

**RESULTS**

The results show that across all SERPs 13.5% of all fixations were on query terms (QFR), while 86.5% of fixations were on other terms and areas on the page (TFR). It is clear that participants fixated on their query terms less than they fixated on other parts of the SERP. Across all SERPs the average AOI ratio (AOIR) was 4%.

A Friedman test was used to investigate differences in QFR and TFR between tasks and between participants. Significant differences were found between tasks (χ² (2, N = 10) = .026, p < .05) and participants (χ² (2, N = 10) = .009, p < .05). However, post hoc tests revealed significant differences only in the extremes. For AOIR, no significant differences were found between tasks (χ² (2, N = 10) = .402, p > .05) or between participants (χ² (2, N = 10) = .483, p > .05).

**DISCUSSION**

The results indicate that searchers fixate on their query terms; however, other terms and other areas of the SERP also receive attention. This result is consistent with prior studies of log data that suggest the presence of query terms affects scanning and search behavior. These values do not suggest that query terms in the result captions exerted a large influence over visual attention, as suggested in two prior studies (Yue et al., 2010; Clarke et al., 2007).

The findings of Eickhoff et al. (2014) suggest that searchers fixate on non-query term words present on SERPs, looking at words that are then used in reformulating queries. This is one possible explanation for the high TFR values found in this study.

It must also be noted that we made no distinction between any other fixation points. It may be that when the rest of the SERP is broken down to a word level and those findings are analyzed, that the 13.5% QFR value will be high in comparison to fixation rates on other types of words and areas of the SERP. This analysis is planned as future work.

**LIMITATIONS**

The study reported here has several limitations that may affect its generalizability. First, participants were drawn from a potentially homogenous pool of university students, all of whom were registered in a core introductory college course and were close in age. Because of time constraints, analysis for only ten participants was performed. Different patterns of scanning and fixation may be detected in a larger and more heterogeneous sample. The first author, who was aware of the purpose of the study, performed all of the AOI identification by hand. An automated method of detecting and coding words on SERPs, such as the process used by Cutrell and Guan (2007), would allow for deeper investigation of the types of words searchers look at on results pages.

**Table 2. A table of all terms exluded as AOIs.**

<table>
<thead>
<tr>
<th>Excluded Terms</th>
<th>a/an</th>
<th>all</th>
<th>and</th>
<th>are</th>
</tr>
</thead>
<tbody>
<tr>
<td>as</td>
<td>at</td>
<td>can</td>
<td>do</td>
<td></td>
</tr>
<tr>
<td>for</td>
<td>from</td>
<td>how</td>
<td>in</td>
<td></td>
</tr>
<tr>
<td>is</td>
<td>of</td>
<td>on</td>
<td>that</td>
<td></td>
</tr>
<tr>
<td>the</td>
<td>their</td>
<td>to</td>
<td>what</td>
<td></td>
</tr>
<tr>
<td>who</td>
<td>why</td>
<td>with</td>
<td>you/r</td>
<td></td>
</tr>
</tbody>
</table>
The first author, who ran all of the data collection sessions, observed that some participants appeared to be focused on finding instances of the nouns from the task description as a strategy for finding a website that contained information on the topic defined by the nouns. Further analysis of the data gathered from this study may reveal whether this was the case. In future research, a different approach will be used in crafting the tasks. For example, we may provide a concept word (herding for the dog and sheep example) and ask participants to find information on that concept without using the single prompt word as a query term.

In the analysis, the Tobii Fixation Filter was used to determine fixations; it does not use a definite millisecond threshold when determining fixations. Other eye-tracking studies have used a 100ms fixation duration threshold when examining eye gaze data (Cutrell & Guan, 2007; Kules & Xie, 2011). While using the threshold would result in different values of QFR and TFR, those differences would not significantly affect the results of this study; only two percent of all SERP fixations had a duration less than 100ms.

CONCLUSION
The goal of this research was to answer the question, “Do people fixate on their own query terms on a search engine results page?” Because previous search studies indicated that participants take their search terms directly from the search task, this work took a novel approach to creating tasks that required participants to come up with search terms on their own. Tobii Studio along with the Tobii T60 eye-tracker were used to gather eye gaze data. When the total number of fixations on each SERP was compared to the number of fixations on query terms, results showed that the participants fixated on their query terms, but less often than other words and parts of the SERP. Understanding where people look on a SERP is important to understanding how people use results pages, particularly as it relates to query reformulation. A better understanding of the way people use SERPs will help in the development of systems that better serve users.

REFERENCES


