

**THE INFLUENCE OF TASK AND GENDER ON SEARCH AND EVALUATION  
BEHAVIOR USING GOOGLE**

Lori Lorigo <sup>a*</sup> lal2@cornell.edu	Bing Pan <sup>a</sup> bp58@cornell.edu	Helene Hembrooke <sup>a</sup> hah4@cornell.edu
Thorsten Joachims <sup>b</sup> tj@cs.cornell.edu	Laura Granka <sup>c</sup> granka@stanford.edu	Geri Gay <sup>a</sup> gkg1@cornell.edu

<sup>a</sup> Cornell University Information Science  
301 College Ave  
Ithaca, NY 14850, USA

<sup>b</sup> Department of Computer Science  
Cornell University  
4130 Upson Hall, Ithaca, NY 14850, USA

<sup>c</sup> Department of Communications  
Stanford University  
Stanford, CA 94305, USA

\*Corresponding Author

July 18, 2005

Submitted for Publication in  
*Information Processing and Management*

# **THE INFLUENCE OF TASK AND GENDER ON INFORMATION SEEKING BEHAVIOR USING GOOGLE**

## **Abstract**

To improve search engine effectiveness, we have observed an increased interest in gathering additional feedback about users' information needs that goes beyond the queries they type in. Adaptive search engines use explicit and implicit feedback indicators to model users or search tasks. In order to create appropriate models, it is essential to understand how users interact with search engines, including the determining factors of their actions. Using eye tracking, we extend this understanding by analyzing the sequences and patterns with which users evaluate query result abstracts returned to them when using Google. We find that the query result abstracts are viewed in the order of their ranking in only about one fifth of the cases, and only an average of about three abstracts per result page are viewed at all. We also compare search behavior variability with respect to different classes of users and different classes of search tasks to reveal whether user models or task models may be greater predictors of behavior. We discover that gender and task significantly influence different kinds of search behaviors discussed here. The results are suggestive of improvements to query-based search interface designs with respect to both their use of space and workflow.

Keywords: Web Search, Eye Tracking, Information Retrieval, Human-Computer Interaction, Information Seeking

## **1. Introduction**

Jansen and Pooch (2000) describe marketing reports stating that web search is the most important Internet activity to users and that one in 28 web pages viewed is a search results page. It is not hard to imagine that the abundance of these interactions may have grown today, together with the increase in material on the web. Because of the inherent value of information discovery, new insights and improvements to Internet search design will have a tremendous impact.

To improve search engine effectiveness, we have observed an increased interest in gathering additional feedback about users' information needs that goes beyond the queries they type in. Adaptive search engines use implicit and explicit feedback indicators to model users (Joachims 2002; Holland, Ester & Keißling, 2003) or to model search tasks, such as those studies presented in the TREC competition (Kang & Kim, 2004; Baeza & Ribeiro, 1999). To create appropriate models, it is essential to understand users' search behaviors.

Earlier work has made great strides in understanding the use of search engines, analyzing the log files of popular search sites including AltaVista (Silverstein, Henzinger, Marais, & Moricz, 1998), Excite (Janson & Pooch, 2000), and also search engines in Europe (Jansen & Spink 2004). These studies reported many descriptive statistics about users' search queries. Looking beyond search engine logs, Granka, Joachims, and Gay (2004) reported initial eye fixation statistics from searches on Google. Recent work by Kloeckner, Wirschum and Jameson (2004) revealed both breadth-first and depth-first sequence patterns on Google. Depth-first searches in this context refer to searches in which a user visits the page of an abstract on the Google result list before reading the

next abstract. Breadth-first patterns occur when a user reads each abstract in the Google result page before visiting the page(s) of interest. Prior to the existence of web search, but perhaps similar in part, MacGregor, Lee and Lam (1986) discovered specific search patterns on menus, which they named terminating, exhaustive and redundant. These prior studies, however, did not address search behavior variability due to the kind of web search task. We wish to not only uncover search strategies but also determine how they vary under different conditions.

To understand the different uses of search engines, Broder (2002) developed a web search taxonomy that classifies the “need behind the query” into three classes: *navigational* tasks are tasks where the user’s intent is to find a particular web page, such as a homepage; *informational* tasks arise when the intent is to find information about a topic that may reside on one or more web pages; and *transactional* search tasks reflect the desire of the user to perform an action, such as an online purchase. Broder’s taxonomy has been widely adopted, as in the work by Kang and Kim (2004), which showed that optimizing search engines based on implicit data about informational vs. navigational search improved performance. His taxonomy was re-evaluated by Rose & Levinson (2004), resulting in rather similar findings regarding the breakdown of these tasks. Broder reported 48% of web queries as informational, and 20 and 30% of queries as navigational and transactional respectively, while Rose & Levinson reported 60% informational, with roughly 14% and 25% navigational and transactional.

While these classifications help explain the different kinds of web search tasks, it remains to be understood if and how these tasks differ cognitively, and if they elicit different information seeking strategies. A strong understanding of the determinants of

Internet search strategies and the appropriateness of interface designs is increasingly vital given the saturation of digital information available. This research analyzes users' eye movement sequence patterns during interaction with Google, and considers how these behaviors are influenced by gender and by task classification. For example, how does effort spent reading abstracts compare with selection behaviors, and how does the effort vary with respect to user and task? Or, can we safely assume that when a user clicks on the  $n^{\text{th}}$  abstract, that she is making an informed decision based on  $n-1$  abstracts preceding it? We present results from our eye tracking study to address these and related questions.

Earlier studies have examined the effect of gender on information-seeking behavior on the web. Large, Beheshti, and Rahman (2002) and Roy and Chi (2003) have both found significant behavioral gender differences in their studies. Large et al. reported on collaborative search behavior, finding teams of boys to be generally more active online. Roy et. al. compared behavioral patterns to learning outcomes, noting that search pattern itself was a strong determinant of learning gains. We utilize gender as one classification of users, so that we may compare its influence on web search behavior against task-specific influences.

Our work extends earlier work in two ways. First, we describe how users evaluate the results returned to them in response to Google queries. Examining this search *within the search results* promotes a richer understanding of the *process* of information seeking (Fidel, 1987). Second, we compare behavior variability amongst users and amongst search tasks, in an effort to understand how user differences and task differences influence search behavior. We use gender to classify users and Broder's (2002) informational and navigational task classes to classify tasks. This extension allows us to

report on the effectiveness and implications of identifying user groups and tasks kinds through behavioral data, applicable when designing customized or adaptive search engines.

In order to better capture the information seeking process, we use eye tracking to supplement traditional log file analysis. For example, if a user selects a page from a list of ranked query results, the click and rank is recorded in a log file, but without also capturing what has been viewed, one cannot report that the user found the earlier abstracts to be inferior. It can help us to understand if the search result page interface design is optimal, and for which uses. Eye tracking has been used in many domains to capture visual cognition as well as user interaction behavior with computers (Rayner 2002; Jacob & Karn 2003).

The remainder of this article is organized as follows. We describe the details of our experiment and our scanpath analysis methods in Sections 2 and 3 respectively. Our results in terms of both overall behavioral indices and also gender and task variations are included in Section 4. We conclude with a summarization and implications of our findings in Section 5.

## **2. Eye Tracking Study**

Thirty-six undergraduate subjects at a large university in the northeast U.S. were recruited for this study. We obtained complete eye tracking data of 23 subjects because of inability of some subjects to be calibrated and additional equipment and setup errors. In total there were 14 males and 9 females. While a smaller sample, this sample remained representative and was subject to thorough analysis (more than 400 queries and 600

Google results pages were studied). The subjects were from several different disciplines including engineering, communication, life sciences, and humanities with an age range from 18 to 23 and were well experienced with the Google search engine, as reported in our survey. For instance, on a scale of 1 to 10, subjects reported their expertise with Google to be 8.8 on average, with only 2 subjects (one male and one female) reporting expertise of less than 7.

----- **Insert Table 1 Here** -----

The five navigational and five informational questions shown in Table 1 were administered to all subjects. All of these questions are closed questions (as opposed to open ended) allowing us to measure success explicitly. Before selecting the specific search tasks to be used in this study, tasks were crosschecked with the top searches listed on the "Google Zeitgeist" ([www.google.com/press/zeitgeist.html](http://www.google.com/press/zeitgeist.html)). The purpose of this was to ensure that the tasks were representative of the various genres of searches that the general population uses on a regular basis, including travel, movies, current events, celebrities, and local issues. The tasks were also pre-tested to ensure that the most intuitive queries would not always result in top-ranked results, and that hence a mixed level of difficulty was targeted across the search tasks. Each subject was surveyed on both the subject's perceived difficulty and prior knowledge of each question, using a scale of 1 to 10. No correlation was found between gender and subjects' reported prior knowledge, or between task and prior knowledge for each question. Nor was a correlation found for gender or task with respect to perceived difficulty. For instance, the most difficult search tasks included both navigational and informational queries, and the average reported difficulty of the questions ranged between 3.06 and 8.06. Two minutes

were given for answering each question. Furthermore, subjects were read the questions aloud in an effort to minimize the tendency to choose only the words in the written questions for their queries and to eliminate unnecessary eye movements away from the computer monitor that could potentially hinder the accuracy of the ocular calibration. A computer script was generated so that each subject performed the ten tasks in a completely randomized order to eliminate unnecessary biases. Subjects were asked to search as usual and did not know what exactly we were measuring. Due to the size of the computer monitor, scrolling was required to view abstracts ranked 7-8 and higher on the Google results pages.

Both transaction log file data and eye movement data were gathered in our experiment. Eye movements of subjects were captured using the Applied Science Laboratories (ASL) 504 eye tracker. The eye tracker captures eye position and pupil dilation with a rate of 60Hz in a natural setting. ASL's Gazetracker software was run to capture the areas of interest on the Google result pages. Areas of interest for our experiment were the 10 regions representing the 10 abstracts of the ranked documents on Google result pages. Subclasses of these regions, such as titlebar, text snippet, and URL address are of interest, but cannot be captured with high accuracy due to limitations of precision of the ASL 504 eye tracker.

We classified eye movements according to the following significant indicators of ocular behavior, namely fixations, saccades, pupil dilation, and scanpaths (Rayner, 1998). A fixation is 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 most information acquisition and

processing occurs (Rayner 1998). Saccades are the continuous and rapid movements of eye gazes between fixation points. Because saccadic eye movements are extremely rapid, within 40-50 milliseconds, very limited information processing happens during saccades. *Scanpaths*, or sequences of fixations, were calculated to represent the eye movements and sequential behavior.

These eye tracking metrics allow us to understand users' interactions to a micro-level. While there are many potential factors that may influence these behaviors, we chose to classify users according to gender and tasks according to informational and navigational tasks as described. These classifications were not only clearly separable, but also served as promising examples with which to illustrate new insights and influence on behavior. To compare differences between genders and between search tasks with we used SPSS to calculate analysis of variance (ANOVA) for various quantifiable metrics of behavior, and Chi square for categorical metrics.

### 3. Scanpath Analysis Methods

While eye tracking can provide a much richer understanding of a user's information seeking strategies, the amount of data can be overwhelming and complex to reason about, particularly when studying scanpaths. There are significant challenges to appropriately extracting scanpath trends; for example, there are multiple ways to measure similarity between two paths, each yielding different results. To help us to reason about search patterns, we define additional properties on scanpaths.

----- **Insert Figure 1 Here** -----

Figure 2 depicts an example *scanpath*, or sequence of eye fixations, on a Google

results page. For analysis purposes, we represent the scanpath in terms of the ordered sequence of Google result abstracts fixated upon by a subject. In the example above, the scanpath is thus  $2 \rightarrow 2 \rightarrow 3 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 1$ . Next we define two abstractions of this path. We obtain the *compressed* sequence by aggregating subsequent fixations that remain on the same abstract into 1 element. Thus, the compressed sequence in the same example is  $2 \rightarrow 3 \rightarrow 2 \rightarrow 1$ . Though we care about time spent viewing each abstract, we are not concerned with fixation travel patterns within a particular abstract or components of the abstract because of the lack of significant precision of the eye tracker and so they are compressed. Our second abstraction of the scanpath is the *minimal* scanpath which is the path obtained by removing repeat visits, or regressions from the compressed sequence. In our example above, the minimal path is simply  $2 \rightarrow 3 \rightarrow 1$ . This representation gives us an overall ordering of the abstracts viewed.

Using these scanpaths, we can better characterize a user's search strategy, and in particular, their search within Google results. We introduce three additional terms to describe this search. First, we say that a scanpath preceding a web document selection (or click) is *complete* if the path contains all abstracts of rank  $n$ , for all  $n$  less than or equal to the rank of the selected web document. Thus the user who makes a complete scanpath does not click on a web page before viewing each of the abstracts that preceded it in the Google ranking. Next, we define a user's evaluation of the search results to be *linear* if the minimal sequence of the user's scanpath is monotonically increasing in steps of 1 and *strictly linear* if the corresponding compressed sequence is monotonically increasing in steps of 1. Note that a compressed path is monotonically increasing if and only if a user makes no skips (jumps of length greater than one) and no regressions (jumps back) while

viewing the abstract list. We also quantify skip and regression distance by the distance of the jumps in number of abstracts between them.

## **4. Results**

In this section, we report on the major findings of this research. We first describe the overall web search behaviors using eye tracking methods, and second describe variations in those behaviors based on task kind and gender.

### **4.1 Search Behaviors**

To give a rich understanding of the interaction, we highlight conclusive results with respect to the search question, individual queries, and individual query result page viewings. Overall about 60% of the questions were answered correctly within the time limit allowed. Over each successfully answered question, on average 1.6 queries were entered to Google, and an average of 4.78 abstracts were clicked upon (including repeat selections of abstracts).

With respect to the individual queries, 437 queries were submitted by the subjects in total. On average 0.80 web documents were selected, or clicked upon per query, and the number of clicks ranged from zero to five. Thirty-nine percent of the queries resulted in at least one abstract selection. In 96% of the queries, subjects looked only at the first Google result page, containing only the first ten abstracts. No subjects went beyond the third results page. Most of the subjects viewed the first and second results rather equally in frequency; however, they chose to click on the first result most of the time. The results indicate that for each query only the most highly ranked search results are likely to be exposed to the users. We also observed that users rely heavily on the abstracts alone to determine the relevance of the corresponding web page. On more than half of the search

sessions, users chose to revise the query terms without clicking on any abstract.

To obtain the patterns of evaluation of the abstracts on the query results pages, we calculated scanpath patterns on over 600 Google result pages. The average lengths of the scanpath, compressed scanpath, and minimal scanpath were 16, 5.8, and 3.2 respectively. Thus subjects viewed on average only 3 abstracts per result page. Sixty-seven percent of sequences for which a document was selected were *complete*. 34% of the scanpaths were *linear* and 19% were *strictly linear*, the latter mostly attributed to sequences of lengths one and two (queries with length one are by default strictly linear). 59% and 50% of sequences have regressions and skips, respectively. For each pair of sequential abstracts in the scanpaths, the average distance between their ranks was 1.67, calculated on over 5000 pairs of eye movements. These findings demonstrate that users do not generally follow the ranking presented by the search engine.

On those Google result pages in which a document was not selected, only 4% of the scanpaths contained all 10 abstracts. This indicates that the user in most cases quickly determines that she will reformulate a query without checking each of the returned abstracts. Not only are the latter ranked result pages rarely clicked upon, but they are also infrequently viewed. This property may suggest that fewer abstracts could better serve users' search behaviors. In general we find that users are looking at the first two results, but if these are not believed to be relevant, they are making skips, jumps and regressions in viewing the abstracts. During this jumping period, users often revert their gaze to reevaluate the first two abstracts again, fixating on average on about six result abstracts including repeat abstracts, or on three unique result abstracts before leaving the results page.

The overall picture shows us that subjects do few queries, and view the first few ranked abstracts primarily, and are often not reading the ranked abstracts in the order they are presented. We next consider how these general trends vary with respect to gender and task.

#### **4.2 User and Task Influence**

Results from an ANOVA test with task kind as a factor and several dependent variables are depicted in Table 2. The variables listed in this table were computed per question. Later in this section we will see additional variables with respect to the individual queries and results abstracts. Only variables that resulted in a significant difference in means for either gender or task kind are displayed. Variables with no significant difference include: rate of success, number of queries per question, number of fixations on Google results pages, time spent on the Google results pages, number of results pages viewed, and average fixation duration. In this table, the time variables were analyzed on successfully answered questions only. Success rates were not significantly different with respect to task or gender. Similar relationships between informational and navigational tasks were achieved when we evaluated these variables with respect to individual queries.

----- **Insert Table 2 Here** -----

In the table, “Percent Time on Google Result Pages” gives the proportion of time spent evaluating abstracts over the overall time spent per question. This distinction was made for measuring interaction with the search engine interface alone versus interaction with web sites outside of Google. Intuitively, a greater proportion of time is spent on

Google result pages for navigational tasks since navigational questions do not require much additional scrutiny on web documents outside of Google; once the appropriate query result is found, one immediately has the answer for the navigational query. Thus, for navigational searches, the results page is typically where the user's need is met. Instead, for informational searches, the search on the respective web documents is a critical part of the search process, and the results page is only an intermediary for that web document search.

Another observed variable influenced by task type is pupil dilation on selected web documents. Pupil dilation has been shown to indicate arousal (Anderson & Pietrzak, 1967; Beatty, 1982) and the significantly greater dilation for informational searches indicates that users need to exert more cognitive efforts on the clicked or chosen query results pages. This is not surprising, but is noteworthy because it offers support that a different cognitive activity is taking place. Note however, that pupil dilation was not significantly different between task types on the Google results pages themselves; it only differed once on a selected page. This implies that the process of searching within the query results abstracts may yield similar levels of cognitive arousal for the two tasks.

Gender had almost no impact on each of the variables listed thus far. The only significant gender findings discovered per question relate to information seeking on selected web documents beyond the Google interface. Specifically, males had greater average fixation durations on web documents than females ( $p < .05$ ). Also, gender differences did not play a role for any of the behaviors listed above at the query level. One additional query specific variable was observed to be significantly different however; females submitted significantly longer queries than men, with means of 3.8 and

3.4 words respectively ( $p < .02$ ).

While task type has had a much greater influence than gender on most of the interactions discussed, the reverse is the case when we consider the evaluation of Google result page abstracts. Results from ANOVA and Pearson Chi square tests using gender as a factor and several dependent scanpath properties are shown in Table 3. The first two variables in the table are relevant only on pages where a document was selected (clicked). Only variables between which a significant difference existed are shown. Variables which were tested and showed no significant influence by gender include time durations, number of fixations, average fixation duration, average pupil dilation, length of the scanpaths, as well as fixations and clicks on the abstract numbers that are not mentioned in the table. The percent of repeat viewings of abstracts and the linearity of the scanpath were compared only on the first viewing of the result page.

----- **Insert Table 3 Here** -----

When considering the number of fixations on each of ranks 7 through 10 and also simply whether or not the abstract for that rank was gazed upon, we observe that the mean values for males are significantly higher. Likewise, males also were 5.4 times more likely to view additional Google results pages than females.

We also observe that females make more regressions, or more often return to previously visited abstracts in their scanpaths with respect to their scanpath lengths, while men's scanpaths are more likely to be strictly linear. Also, disregarding repeats, the overall order in which the abstracts are viewed is more linear in men's scanpaths. The lengths of the paths were not significantly different, however. These findings suggest that while Google's linear presentation of results may be ideal for some strategies, others may

benefit from some forms of record keeping of evaluations (such as check boxes, for example) or nonlinear presentations.

None of the above values were significantly different between informational and navigational tasks. The only significant task findings discovered were that (1) the number of abstracts viewed below the clicked rank was greater for navigational searches, and (2) fixating on rank two was more likely for informational than navigational tasks, significant at .03.

Since in many of the cases only two abstracts are viewed, simplifying the evaluation, we also looked solely at the data for which more than two abstracts were viewed, representing cases where the first two abstracts were likely not obviously the solution to the search task. For these 327 scanpaths, the results are even stronger. That is, each of the variables noted in Table 3 remained significant and an equal or greater significance, plus additional behaviors differed between genders. Also, no significant differences were found in informational vs. navigational tests. The additional differences between genders represent the finding that males look at a greater number of abstracts and also spend more time examining the result page in general. Whether or not abstract 6 was viewed also became significantly higher for males. This is worth mentioning since it is suggestive that the greater tendency to view abstracts 7 through 10 by males is not solely influenced by the need to scroll to see abstracts 7 or 8 and higher.

## **5. Conclusion**

We have presented statistics about web search activity in an effort to deepen our understanding of users' interactions with query result pages in the search process, and to

show ways in which task and user classes can influence search behavior. Our study captures the behaviors of a sample of college students on a series of 10 closed search questions. We find that in general time and pupil dilation are influenced by whether the search task is informational or navigational and not by gender, while patterns for evaluating the query result abstracts are influenced by gender and not by task kind. Informational searches took more effort and time on average, despite no task influence found with respect to success. Males looked at abstracts further in the result list, were more linear in the order in which they viewed the abstracts, and made fewer regressions. The sensitivity to rank position in our results is significant to both interface designers and information retrieval researchers, particularly because search engine ranking functions can produce nearly equivalent weights for the top five or ten ranked documents.

Eye tracking methods and particularly scanpath analysis provided a level of detail unreported in previous web search studies. It is valuable to consider web search behavior on multiple levels because both macro and micro level behavioral properties are important for design. By presenting several factors of influence, and comparing findings according to user and task types, we observe that task and gender each influence different kinds of search behaviors, with evaluation patterns more sensitive to gender and cumulative question data more sensitive to task. Direct implications of gender differences and their grounding in gender studies theories are beyond the scope of this work.

In the future, it would be of enriching to analyze search and evaluation behaviors in terms of information foraging theory, developed by Pirolli and Card (1999), or other earlier information seeking models, in an effort to compare the cost, or perceived cost, of reformulating a query, the cost of clicking on an suggested document link, and the cost of

gazing, or reading the abstract, over time. We also have initial findings on average sequences through the abstracts or the selected documents in an effort to put together a more comprehensive visualization of search patterns. Additionally, there has been a considerable amount of work on the design and evaluation of web search systems to incorporate both direct search and browsing activities (Hearst, English, Sinha, Swearingen & Yee, 2002). Drawing on those results and doing further eye tracking studies of alternative search systems could be fruitful.

### **Acknowledgements**

The research was partly funded by Google, Inc. Thanks also to Mathew Feusner and the Human-Computer Interaction Group of Cornell University for valuable discussions, and to the reviewers for their valuable suggestions.

### **References**

- Anderson, R. O., Pietrzak, P. E., (1967). Pupil dilation as a measure of workload (Pupillometric experiments to measure work capacity and task complexity) NASA 3D ANN. NASA UNIV. CONF. ON MANUAL CONTROL 1967 /SEE N68-15901 06-05/ P 305-308.
- Baeza, R. & Ribeiro, B. (1999). Modern Information Retrieval. Addison-Wesley, 1999.
- Beatty (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychological Bulletin, 91, 276-292.
- Broder, A. (2002). A Taxonomy of web search. SIGIR Forum, 36(2): 3-10.
- Byrne, M. D., John, B. E., Wehrle, N. S., & Crow, D. C. The tangled web we wove: A

taskonomy of WWW use. Proceedings of CHI, 1999 (Pittsburgh, PA, May 15-20, 1999) ACM, New York. pp. 544-551.

Fidel, R. (1987). Database Design for Information Retrieval: A Conceptual Approach. New York: Wiley.

Granka, L., Joachims, T., & Gay, G. (2004). *Eye-Tracking Analysis of User Behavior in WWW-Search*, Poster Abstract, Proceedings of the Conference on R&D in Information Retrieval (SIGIR), 2004.

Hearst, M., Elliott, A., English, J., Sinha, R., Swearingen, K., and Yee, P. (2002). Finding the Flow in Web Site Search, Communications of the ACM, 45 (9), September 2002, pp.42-49.

Holland, S., Ester, M., & Kießling, W. (2003). Preference Mining: A Novel Approach on Mining User Preferences for Personalized Applications, Knowledge Discovery in Databases: 7th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), 2003.

Jacob, R.J.K. & Karn, K.S. (2003). Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises (Section Commentary), in *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, ed. by J. Hyona, R. Radach, and H. Deubel, pp. 573-605, Amsterdam, Elsevier Science, 2003.

Jansen, B.J., & Pooch, U. (2000). Web User Studies: A review and framework for future work. *Journal of the American Society of Information Science and Technology*, 52(3), pp. 235-246.

Jansen, B.J., & Spink, A. (in press). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing and*

Management.

Joachims, T. (2002). Optimizing Search Engines Using Clickthrough Data, Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD).

Kang, I. & Kim, G. (2004). Integration of multiple evidences based on a query type for web search, *Information Processing and Management: an International Journal*, v.40 n.3, p.459-478.

Kloeckner, K., Wirschum, N. & Jameson, A. (2004). Depth- and Breadth-First Processing of Search Result Lists, Poster Session, CHI 2004.

Large, A., Beheshti, J., & Rahman, T. (2002). Gender differences in collaborative Web searching behavior: an elementary school study, *Information Processing & Management*, Volume 38, Issue 3, May 2002, Pages 427-443.

MacGregor, J.N., Lee, E.S., & Lam, N. "Optimizing the structure of menu indexes: a decision model of menu search," *Human Factors*, 1986, 28, 387-400.

Pirolli, P.& Card, S. (1999). Information foraging. *Psychological Review* 106(4): 643-675.

Rayner, K. (1998). Eye movements and information processing: 20 years of research. *Psychological Bulletin* 124(3): 372-422. Baeza, R. & Ribeiro, B.. *Modern Information Retrieval*. Addison-Wesley., 1999.

Rose, D. E. & Levinson, D. (2004). Understanding User Goals in Web Search, WWW 2004, New York, USA, May 2004.

Roy, M., & Chi, M. T. H. (2003). Gender differences in patterns of searching the Web. *Journal of Educational Computing Research*, 29, 335-348.

Russo, E.J. & LeClerc, F. (1994). An Eye-Fixation Analysis of Choice Processes for Consumer Nondurables. *Journal of Consumer Research*, 21 (2), 274-290.

Shneiderman, B., Byrd, D. & Croft, W.B. Sorting Out Searching A User-Interface Framework for Text Searches. *Communications of the ACM*. 41(4): 95-98.

Silverstein, C., Henzinger, M., Marais, J. & Moricz, M. (1998). Analysis of a very large Alta Vista query log. Technical Report 1998-014, COMPAQ Systems Research Center, Palo Alto, Ca, USA, 1998.

Smyth, B., Balfe, E., Briggs, P., Coyle, M. & Freyne, J. (2003). Collaborative Web Search. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence, IJCAI-03*, Morgan Kaufmann (2003) 1417-1419 Acapulco, Mexico.

<http://www.google.com/press/zeitgeist.html>.

Table 1: Navigational and Informational Tasks.

---

Navigational Tasks
<i>Find the homepage of Emeril – the chef who has a TV cooking program.</i>
<i>Find the page displaying the routemap for Greyhound buses.</i>
<i>Find the homepage of Michael Jordan, the statistician.</i>
<i>Find the homepage for graduate housing at Carnegie Mellon University.</i>
<i>Find the homepage of the 1,000 Acres Dude Ranch.</i>

---

Informational Tasks
<i>Who discovered the first modern antibiotic?</i>
<i>What actor starred as the main character in the original 'Time Machine' movie?</i>
<i>You are excited to cast your vote in the democratic presidential primary – when can you do so in NY?</i>
<i>Where is the tallest mountain in NY located?</i>
<i>A friend told you that Mr. Cornell used to live close to campus- between University and Stewart Aves. – does anyone live in his house now; if so, who?</i>

---

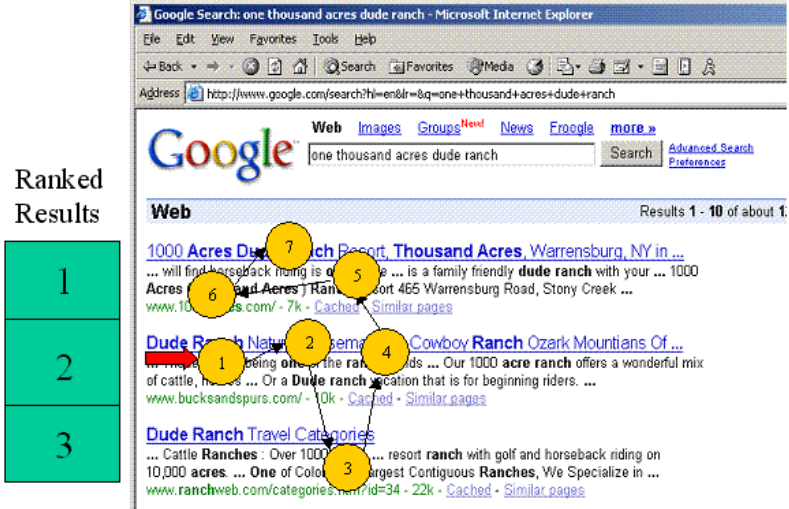


Figure 1. An example of a scanpath on a Google results page.

Table 2: Comparison of variance between informational (I) and navigational (N) search questions.

Dependent Variable	Relationship	Means for Informational Task (I)	Means for Navigational Task (N)
Average time to complete the task	I > N *	46.0 s	34.2 s
Average time spent on web documents per question	I > N*	26.1 s	17.9 s
Percent time on Google result pages per question	N > I *	40.2 %	51.5%
Avg. number of fixations on a web document over documents selected per question	I > N**	23.8	12.8
Avg. time on a web document over all of the web documents selected per question	I > N**	12.1 s	5.8 s
Avg. pupil dilation on a web document over documents selected per question	I > N*	51.6 mm	45.9 mm

\* Significant at .05 level

\*\* Significant at .01 level

Table 3: Comparisons of means between males (M) and females' (F) evaluations of Google result abstracts

Dependent Variable	Relationship	Means for Male (M)	Means for Female (F)
Average rank of selected document	M>F*	1.72	1.59
Percentage of subjects clicked on Abstract 2	F>M**	7.0%	14.5%
Average numbers of result pages viewed	M>F**	1.06	1.01
Percentage of strictly linear scanpath	M>F**	32.0%	20.4%
Percentage of linear scanpath	M>F*	42.8%	33.9%
Percentage of repeat viewings of abstracts	F>M**	26.2%	32.1%
Percentage of subjects fixated on abstract 2	F>M*	53.4%	62.2%
Percentage of subjects fixated on Abstract 3	F>M*	42.7%	51.6%
Percentage of subjects fixated on Abstract 7	M>F**	19.1%	9.5%
Percentage of subjects fixated on Abstract 8	M>F**	18.3%	7.2%
Percentage of subjects fixated on Abstract 9	M>F**	17.8%	6.3%
Percentage of subjects fixated on Abstract 10	M>F**	16.5%	5.4%

\* Significant at .05 level

\*\* Significant at .01 level