Eyetracking in Online Search

Laura Granka[1], Matthew Feusner[2], Lori Lorigo[3]

1. Google Inc., User Experience Research, 1600 Amphitheatre Pkwy,
Mountain View, CA 94043, USA
granka@google.com

2. University of California, San Francisco, Department of Ophthalmology
San Francisco, CA 94143, USA
feusnerm@vision.ucsf.edu

3. Cornell University, Human-Computer Interaction Group
Ithaca, NY 14853, USA
lal2@cornell.edu

Abstract. The Web introduces both new opportunities and challenges for eyetracking research. While eyetracking is still a relatively new analysis tool for studying online human-computer interaction, effective methods for analysis have already emerged. Researchers have used eyetracking to answer a number interesting questions, including online viewing of ads, web homepages, and search results. In web search, eyetracking analysis has illuminated what people decide to look at, how they navigate search results, and what aspects of search pages are the most important for finding information online. Techniques are now available to quantify, compare, and aggregate eye movements relative to these online environments. Unfortunately, while readily available, general eyetracking analysis software has not kept pace with all of the analysis techniques used by researchers. There are certainly obstacles to using eyetracking in online contexts, but tangible results have demonstrated its value and the need for further research.

Keywords: Eyetracking, information retrieval, search

1 Introduction

The Web introduces both new opportunities and challenges for eyetracking research. Due to recent technological advances in hardware and software, the use of eyetracking for the analysis of online applications has rapidly increased. Eyetracking is now being used to offer insights into homepage marketing, advertising, reading of online news, and interpreting user behavior in online search environments. This chapter will assess some of the key issues surrounding data collection and analysis in online contexts. We will then specifically address the context of online search in more detail.

Eyetracking is becoming a popular tool for understanding user behavior in a number of computer and web based contexts, ranging from how viewers read online news stories, [1], how net surfers respond to banner ads and other advertisements [2, 3], and how users interact with web displays and menus [4]. Eyetracking has been used only since 2003 to investigate online information retrieval, which is the context the rest of this chapter will focus [5, 6].

Due to technological advances, we are now able to more effectively answer questions about how a user scans and searches for online content, as well as combine this behavioral data with implicit forms of feedback, such as server log data. We will then discuss how eyetracking is an effective tool to augment standard analysis methods for studying information retrieval.

2. Methodology

2.1 Eyetracking Hardware

Eyetracking lets a researcher understand what a user is looking at while performing a task. There are many different methods for tracking eye movements, including video eye trackers (a type most suited for online and usability contexts), magnetic coil systems (placed directly on a subject's eye that is most used for medical research), and electro-oculography (EOG) recordings (based on muscular movements surrounding the eye). Despite some advantages in recording quality, coils and EOG systems can be very uncomfortable and invasive for the viewer. Video eye trackers can be slower (sampling at about 30-60 Hz) and less accurate due to the need to capture and process each frame, and lose tracking during blinks. However, they are more natural for the participant, and recent technological advances have greatly improved their accuracy.

Ease of calibration also must be considered when selecting eyetracking hardware, especially for use in online applications. If the process is too long or involved, it could be disruptive to a participant's perception of routine web browsing. Various automatic calibration systems are helpful for adding transparency to the use of eye tracking in online contexts, but such options should be considered carefully before trading calibration accuracy for convenience.

Eyetracking experiments traditionally have been very costly and time consuming, though newer technology and improvements in hardware are opening new avenues and research directions. Hardware platforms such as the Tobii 1750 and x50, and ASL R6 VHT [7, 8] are now equipped with automated calibration, which is ideal for industry practitioners who use eyetracking on a regular basis and need a quick and simple set-up.

2.2 Interpreting Eye Movements

Since its emergence as a popular research method, much progress has been made towards accurately interpreting the different eye movements that are captured by eyetrackers. There are now commonly accepted interpretation standards for ocular indices. A careful understanding of these is necessary to accurately interpret the collected eyetracking data. Several key variables have emerged as significant indicators of ocular behaviors, namely fixations, saccades, pupil dilation, and scanpaths [9, 10].

Fixations/ fixation duration. Though there are many different approaches to identifying fixations [11], a fixation is generally 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 traditionally are understood to be indicative of where a viewer's attention is directed, and represent the instances in which information acquisition and processing are able to occur [9]. Based on existing literature, a very high correlation has been found between the display item being fixated on and that being thought about. Similarly, there is a close connection between the amount of time spent fixating on certain items and the degree of cognitive processing [12, 9]. Eye fixations are the most informative metric for evaluating information processing primarily because other indices, such as saccades, occur too quickly for the viewer to absorb new information [9].

At least three processes occur during an eye fixation: encoding of a visual stimulus, sampling of the peripheral field, and planning for the next saccade [13]. Research has shown that information complexity, task complexity, and familiarity of visual display will influence fixation duration [14]. The length of eye fixations is also largely dependent on a user's task. The average fixation duration during silent reading is approximately 225 milliseconds, while other tasks, including typing, scene perception, and music reading approach averages of 300-400 milliseconds.

From an eye tracking perspective, information retrieval seems to encompass both a visual search scenario as well as reading, so it is expected that the average fixation duration will fall within the range of these two groups. The differences in fixation length can be attributed to the time required to absorb necessary information, and the speed at which new information should be absorbed. It is necessary for the eye to move rapidly during reading, while in visual search and scene viewing, it is less imperative that the eye quickly scans the entire scene, but rather that the user can absorb key information from certain regions.

Saccades. Saccades are the continuous and rapid movements of eye gazes between fixation points. They are extremely rapid, often only 40-50 milliseconds, and can have velocities approaching 500 degrees per second. No information is acquired by the viewer during a saccade due to the unstable image on the retina during eye movements and other biological factors. This lapse of information intake is traditionally referred to as "saccadic suppression," but because saccades represent such short time intervals, individuals are unaware of these breaks in information perception [9]. Saccadic movement has been analyzed extensively in the context of reading. The research has identified saccadic behaviors including regressions (re-reading content) and word skipping, which are also important to consider in online contexts [9].

Pupil Dilation. Pupil dilation is a measure that is typically used to indicate an individual's arousal or interest in the viewed content matter, with a larger diameter reflecting greater arousal [14, 9, 15]. Studies can compare the average pupil dilation that occurs in a specific area of interest (AOI) with the average pupil dilation of the entire site to gain insight into how users might cognitively understand or process the various content matter [16].

Scanpath. A scanpath encompasses the entire sequence of fixations and saccades, which can present the pattern of eye movement across the visual scene. User scanpath behavior provides insight into how a user navigates through visual content. Studies analyzing properties specific to scanpath movement have enabled researchers to create a more comprehensive understanding of the entire behavioral processes during a visual or online search session [17, 18, 19]. Existing literature suggests that scanpath movement is not random, but is highly related to a viewer’s frame of mind, expectations, and purpose [20]. Several researchers have explored the sequence of eye movements more closely using sequencing alignment algorithms, which will be discussed further on in the chapter [21, 19].

Other Eye Movements. Smooth pursuit is a type of eye movement where a fixating eye smoothly drifts in order to follow a moving target [22]. This type of viewing does not produce saccades, and generally is not analyzed in online eye tracking for two reasons. First, web page content is almost entirely static; text and images generally are not animated. The emergence of embedded videos and Flash animations may prompt further study, but to date studies have focused on uncomplicated static scenes. Secondly, smooth pursuit tracking only occurs due to moving page elements, such as when a user decides to scroll. However, time spent scrolling is assumed to be low compared to total time viewing a static scene, so a user’s time spent in smooth pursuit eye movements is assumed to be negligible and omitted.

Vergence movements can occur when the eyes move inwards or outwards together in order to refocus at a new distance. However, viewing distance is almost always stable during computer use, so vergence movements are also assumed to be negligible.

Area of Interest (AOI). Often, a researcher is interested in analyzing eye movements with respect to specific regions of a scene, or webpage, such as ads, images, and primary content areas. For this purpose, metrics such as number of fixations, fixation duration, or even pupil dilation, are often reported per each area of interest (AOI), also known as a "lookzone" in some software applications. Classifying unique regions of interest on a page lets a researcher make comparisons between, or even draw conclusions about specific types of content, such as whether there are differences in eye movements when viewing advertising versus standard content. Most eyetracking software used in online applications comes with basic tools to map gaze coordinates to lookzones, though this functionality is often very cumbersome and ill equipped for extensive investigations on large numbers of pages. We discuss several software options that address some of these concerns later in this paper.

2.3 Eyetracking Methodology

Qualitative Analyses. In addition to formal controlled eyetracking experiments, industry practitioners and user experience researchers are now developing qualitative approaches to using eyetracking in day-to-day evaluations of web-based products [26]. In industry work, there is value in simply viewing the pattern and path that a user takes when interacting with a new product or design. For usability studies, Granka and Rodden [26] discussed the benefits of using the eyetracker, not necessarily for the in-depth data analysis, but because of the immense value in enabling product teams to view the real-time projection of the user's eye gaze during the completion of tasks in a usability study.

Qualitative analysis in eyetracking stands in stark contrast to tightly controlled experimental design, where much care must be taken to conduct appropriate statistical analyses. Because eyetracking data is so complex and multi-leveled, to fully account for all random and fixed effects in most eyetracking experiments (e.g., assessing how task type and gender impact fixation duration and pupil dilation), and to generate appropriate estimates of error based on the nested data structure, three- and four-level linear mixed models are often the ideal solution to accurately analyze the data, particularly if the researcher is interested in differences between conditions, task types, and other related metrics [27].

Quantitative Analysis. For quantitative eyetracking analysis, Goldberg and Kotval [23] summarize methods for analyzing eyetracking data for computer-based usability studies, looking specifically at the use of menus and toolbars. Many of their suggested analyses, such as assessing the length of a scanpath, determining the fixation density in a region, and comparing fixation durations, are now regularly used metrics in a lot of recent web-based eyetracking work. These numerical metrics (numbers, durations, and lengths of fixations and saccades) are analyzed using standard statistical methods.

Another type of data analysis that Goldberg and Kotval discuss [23] is fixation clustering. Fixation clustering is particularly important for identifying the division of attention between different segments of a web page. Such analysis can be done visually with a fixation map [24], but also analytically using spatial clustering algorithms [25].

Sequence Alignment Methods. Sequence alignment is an interesting technique for measuring the similarity between two scanpaths over the same stimulus. Scanpaths are interpreted as sequences of fixations, which can be compared using a well-known optimal pairwise alignment algorithm [28, 29]. The magnitude of dissimilarity between the scanpaths is computed by calculating the smallest combination of insertions, deletions, and replacements that need to be made in order to transform one sequence into the other. Each edit (insertion, deletion, substitution) has an associated cost, and the sum of costs is called the edit distance or Levenshtein distance. Josephson and Holmes [17] have used this pairwise matching method to compare scanpaths across different structural elements on a webpage. Another research effort used sequence alignment to group scanning behaviors on popular websites [30]. As an extension to sequence alignment, multiple sequence alignment is now being used to analyze and aggregate populations of viewers [19, 31]. Multiple sequence alignment algorithms generally build on pairwise alignment by repeatedly aligning sequences until one sequence is left that contains elements of all scans in the group.

One unresolved issue regarding both pairwise and multiple sequence alignment algorithms is the choice of algorithm parameters. The first parameters needed for the Levenshtein distance are the insertion, deletion, and substitution costs. By default, unit values are used to equate total cost with total number of edits. However, in the two-dimensional space on a computer monitor, it makes sense to assign a substitution cost according to the distance in pixels. Consequently, insertion and deletion costs must be assigned in the same units. When substituting fixations based on their distance, it makes sense that an insertion or deletion edit would also involve an imaginary associated saccade. Therefore it also would make sense to use the average saccade length in pixels for both the insertion and deletion score.

Probabilistic Models. Probability values computed from eyetracking data can be incorporated into many different descriptive models that represent eye movements. After configuring such a probabilistic model with recorded data like scanpaths, the model can be used to evaluate other scanpaths and eye movements. While use of these models is often more complex than sequence alignment methods [32] it can be more powerful as well. A transition matrix is one simple but effective way to model free-viewing eye movements as in visual search tasks [33]. Probabilities of movement between different locations are computed, and then interesting transitions are selected and evaluated. For example, horizontal, vertical, or diagonal movements between different elements indicate varying levels of effectiveness during reading or scanning tasks.

Markovian models compose one class of probabilistic models that incorporates varying levels of probabilities. For example, a first-order Markovian transition matrix would capture not only the probability of movement from each element to the next, but from the previous element or state as well. Higher order models capture transitions further and further into the past. A Hidden Markov Model can be used when states between transitions are not known or explicit. The complexity captured in the depth of order in a Hidden Markov Model used to model different eye movement strategies can indeed affect the results [34]. However, in print media, it has been shown that even a first-order Markov model can capture a great deal of interesting information about a scanpath [35].

Error Analysis. When conducting research in online environments, it is important to understand the degree of accuracy with which to interpret your findings. Currently the most popular monitor size and resolution is a 17" monitor set to display 1024x768 pixels. At standard ergonomic viewing distances between 24 and 40 inches [36], eyetracking hardware error can be surprisingly large. Many eye trackers claim to be accurate to within 1 degree of visual angle, which corresponds to an on-screen error of between 32 and 53 pixels (table). A single line of text on the web is often around 10 pixels high, so in a normal viewing setting with a popular eye tracker it is unlikely that the researcher can be certain which line of text is being viewed. Some studies have compensated for this problem by using large monitors [37], but the ecological validity is a concern if the text or monitor size is drastically larger than what a user is typically sees.

The distribution of error also is not usually described for popular commercial eye trackers. What fraction of samples is within 0.1 degrees of the true value? In the absence of a complete study, it is worth noting that while one degree of error seems small, it can amount to a large number of pixels, so it is hard to draw conclusions about the specific words a user fixates. It also is worth noting that the number of pixels corresponding to 1 degree of visual angle on a flat monitor begins to increase dramatically around 30 degrees from center. However, even on a 21" monitor at a 24" viewing distance, the screen only spans +/-19.3 degrees of the visual field.

Thus, while very robust, eyetracking is not always the most suitable method for many questions about online analysis. While it can accurately portray a user's typical course of action on a search results page, it is costly and only possible for studies where the user is physically present in front of the equipment. Calibrating and monitoring of the study also add overhead, meaning that analyzing and aggregating all user data from one study requires more work than even multi-session log analyses.

Table 1. Pixels per degree of visual angle for popular monitor sizes, screen resolutions, and viewing distances.

Diagonal Screen Size	24" viewing distance			40" viewing distance		
	1024 x 768	1280 x 1024	1600 x 1200	1024 x 768	1280 x 1024	1600 x 1200
14"	38.98	48.73	60.91	64.24	80.31	100.38
15"	36.48	45.60	57.00	60.02	75.02	93.78
17"	32.37	40.46	50.57	53.07	66.33	82.92
19"	29.14	36.42	45.53	47.59	59.49	74.36
20"	27.77	34.72	43.40	45.27	56.59	70.73
21"	26.54	33.18	41.47	43.17	53.96	67.45

Software and Data Visualization. Software for visualizing and analyzing eyetracking data is either custom-made or packaged with eyetracking hardware. Heatmaps, such as those shown in Figure 2, in the following section, are a popular and effective means to show aggregate viewing behavior on a given web page, and can be found in some of today's latest commercial eyetracking software [38]. Some other commercial options, which are often bundled with hardware, include GazeTracker [39], and iView X [40], and each of these packages vary in the features they provide. Although there are efforts underway to standardize various eye tracking data formats and protocols [41], and there are some freely available analysis programs like WinPhaser [42], the availability of, and functionality afforded by software for analysis should be considered before beginning an eye tracking study.

One key limitation of current commercial eyetracking software is that there is little support for analyzing and discerning patterns in the scanpaths themselves. At present, eyetracking data is typically visualized either as an aggregate view of what was viewed, as in heatmaps, or as individual scanpaths, as in Figures 3 and 4. While both of these techniques offer great insights over traditional log analysis, they lack the ability to convey what an "average" sequence looks like, how typical a given path is compared to another, or where common subsequences lie. One sequence analysis program, eyePatterns, announced its offering for free download [31], but is unavailable as of this writing. Pellacini et al. [21] also describe a work in progress for visualizing paths in the context of a collection of paths, and Hembrooke et al. [19] discuss methods for extracting such an "average" path. Scanpath analysis provides additional information about the process involved in a user's interaction with a SERP, though at present is often too cumbersome due to the scarcity of existing tools and the complex nature of the sequential data. Scanpath analysis is particularly useful in the context of information retrieval, where the path of eye movements offers direct insight into a user's decision making.

3 Eyetracking in Information Retrieval

The greatest strength of eyetracking in the context of information retrieval is its ability to highlight what a user is looking at before selecting a document, depicting the process whereby individuals view the results presented to them on a search SERP. In contrast to traditional methods that study information retrieval behavior, eyetracking is better suited to assess the actual behaviors that users employ when reading and making decisions about which documents to select. Eyetracking will enable us to detect what the searcher is looking at and reading before actually selecting an online document, while other methods for information retrieval analysis, such as server log data, are all based on the single result that a user has selected.

Figure 1 explains the terminology that we will discuss throughout the rest of the chapter. Search engines typically return 10-20 results per page. We refer to each of these shortened results as the surrogate, which is comprised of a title, snippet, and URL. Figure 1 shows a sample Google result page with the key terminology highlighted. Later on in the chapter, we will discuss additional components to the page such as sponsored links, which are also highlighted in the example below.

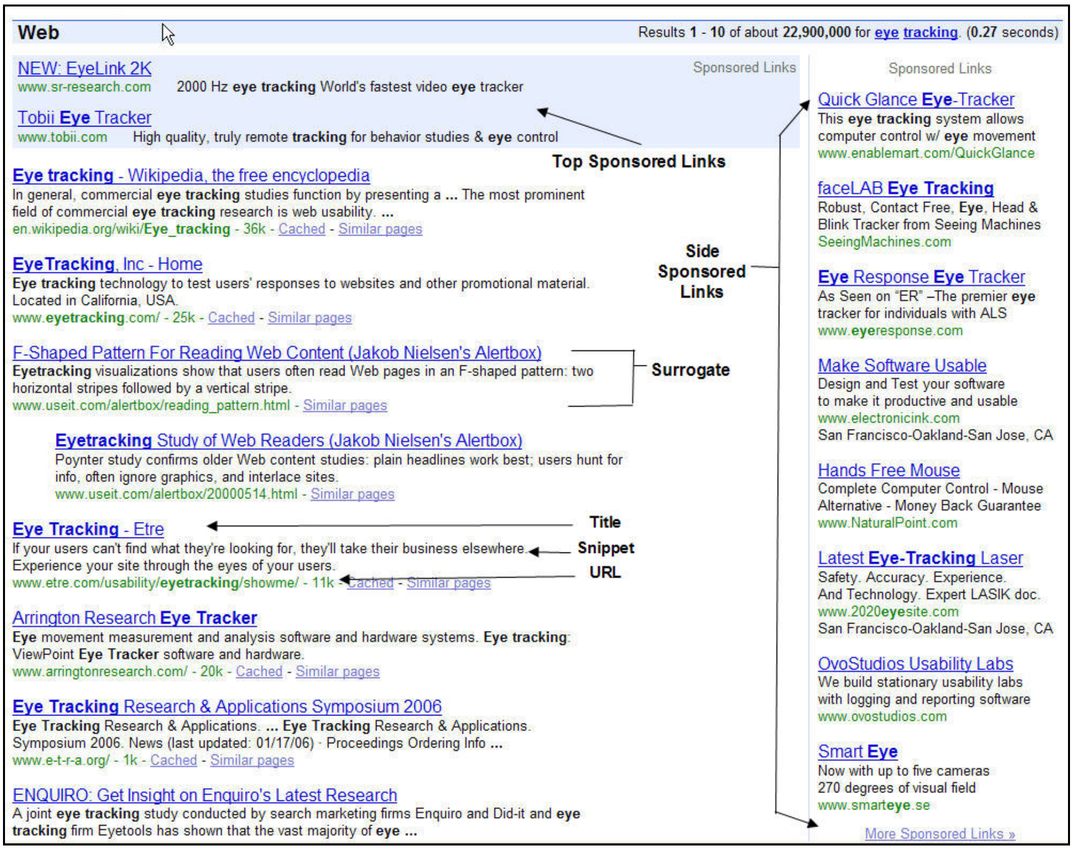

Figure 1. Detailed breakdown of search engine result page (SERP) content.

3.1 Relevant Information Retrieval Research

The information retrieval literature can largely be split into two distinct classes of research – those that investigate the effectiveness and quality of the search engine, and those that evaluate the behaviors that users employ when engaging in an online search. While the eyetracking research discussed here falls into the latter, the findings also provide direct recommendations for search quality, particularly those that use log data to measure result relevance [43, 44].

A majority of current information retrieval evaluations focus on evaluating the functionality of the actual search system, in terms of its efficiency, precision, and ranking of results. These retrieval evaluations focus on assessing the precision and recall of the selected documents as a measure of the information retrieval system itself. Somewhat less focus has been placed on the user - towards addressing the cognitive behaviors involved in evaluating online search results [45, 46, 47]. Our primary objective with eyetracking is to examine how searchers assess, evaluate, and selectively deem an online document to be relevant.

Capturing eye movements during search will address pre-click user behaviors and provide a comprehensive account of how a user views and selects an online document. Analyzing eye movements will enable us to recognize what a searcher reads, skips, or scans, and it is these metrics that offer insight into the use of current information retrieval systems.

Effectively using the ocular indices previously described will enable us to determine the content that searchers focus on, for how long, and in what order, enabling us to examine some of the assumptions that have limited the traditional evaluations of information retrieval systems [48]. As will be discussed in the last section of this paper, relating eye movement behavior with other measures, such as clickstream or mouse movement data, can provide an even more comprehensive picture of the process through which online information acquisition actually occurs.

In sum, the common trend throughout many of the existing user studies is that they look primarily at the outcome measures of user behavior, quantified in such ways as query wording, time spent searching, and the rank of the selected document. These measures are produced after the subject has selected a document. Eyetracking can be used to supplement these metrics, enabling us to understand the actual process whereby users reach a decision about which document to select.

3.2 Existing Research Findings

To date, a number of studies have investigated viewing behavior on search engine result pages (SERPs), the first of which offered a primarily descriptive explanation of how users view the results presented to them [49, 50, 51]. Several researchers have subsequently gone beyond these descriptive measures to also assess what these eye movements mean in the context of interpreting server log data, as well as determining an "average" course of viewing on the SERP [18, 48]. This section will discuss some of the key findings in the eyetracking and information retrieval literature, pointing out some of the slight differences in the data that are produced.

3.3 Overall Viewing Patterns

Several reports have likened the path of a user's eye on a search results page to an F shape, or a Golden Triangle, with the majority of attention being given to the top few results [50, 51]. Figure 2 presents a standard SERP with the golden triangle pattern of viewing. Aggregate analyses show that while users may read the first result in detail, they will rarely give this same degree of attention to the following results, thus each result is read successively less, tapering into the bottom of the triangle. Based on this finding, Jakob Nielsen again stressed the importance of effective web writing, indicating that users won't read your text word-by-word, and the most important information should be stated clearly in the beginning of the sentence or paragraph, ideally with bullet points or "information carrying words" [50].

While the golden triangle, or F-shape is certainly a relevant generalization, work has been done to suggest that the viewing behavior on SERPs is not quite so simple.

In fact, viewing patterns highly depend on the user and task. In the first study produced by Enquiro, the authors called out attention to slight variations to the golden triangle produced by oneboxen and sponsored links [51]. They found that with onebox content and sponsored links present, the top portion of the triangle is elongated to compensate for the additional text.

Recent work goes even further to suggest how tasks and users can impact the way in which a results page is viewed [52]. Images 3, 4, and 5 depict one given task that asked the user to find out who is the tallest active player in the NBA. Users were provided with the query term [tallest active player NBA], and were instructed to begin with that query to find the answer [52]. The heatmap presented in Image 3 is an aggregate of all 32 users who did this task, and shows some slight resemblance to the golden triangle and F-shape. The main similarity is that the first result is read the most completely, the second result is read fairly completely, and the following results have successively fewer fixations.

Figure 2. Golden triangle viewing: the aggregate plot of eye movements can resemble a triangle, with top results viewed more extensively than the lower.

Individual User Differences. Upon closer inspection, however, it appears that users seem to be on average reading the snippets rather carefully, more thoroughly than the "golden triangle" generalization would suggest. An explanation for this behavior is that the answer to this task is explicitly stated in the snippet of the second surrogate. This offers evidence that slight differences in the results page and content presented, such as these information-bearing snippets, have the ability to produce different viewing behaviors. Thus, it is important to understand and account for the different variables on the page when interpreting aggregate plots and making assumptions about user behavior.

To identify individual user behaviors, two individual scan patterns are presented in Images 4 and 5 to emphasize the degree of variance between individuals, even when viewing the same exact page. In the Image 4, the user is quickly scanning only a few results before feeling satisfied and clicking on one. In contrast, the user in Image 5 is evaluating the results much more carefully, reading snippets in depth, as well as exhausting more of the options available to her. Both users were presented with the same page, but approached the information gathering process very differently.

Aula et al. [53] also noticed these individual differences in scan activity in her own research. The authors used eyetracking to assess how users evaluate search result pages, and subsequently classified two types of online searching styles - economic and exhaustive. Economic evaluators are similar to the first image shown, where users scan quickly and make a decision based on what seems to be the first relevant result. Exhaustive evaluators, on the other hand, are fairly similar to the second image depicted below, where the users prefer to assess multiple options before clicking on a result.

Much more work is needed to determine if scan behavior remains consistent on a per-user basis across a variety of tasks (e.g., does a given user always scan SERPs quickly?), or whether the task and presentation of results ultimately have the most impact in determining how the results are viewed. While research has found differences in viewing search results based on variables like task type and task difficulty [18, 49], less user-centric analysis has been done.

Figure 3. Note the golden triangle is stretched a bit. Users are reading the snippets more thoroughly than the golden triangle would suggest.

Figure 4. "Efficient" style of scanning a SERP – viewing less than three results.

Figure 5. This user view the page much more exhaustively, reading all of the surrogates in more detail, and taking the time to look at more than three.

3.4 Number of Results Viewed

While the previous example shows that users exhibit different behaviors in search, it is still helpful to know, on average, how exhaustively users attend to the information presented to them. On average, users only view about three to four surrogates on the SERP [49, 51, 18, 48], and the number seems to vary based on the task and the expertise of the user. The research done by Granka, Joachims, and Lorigo [18], [48], and [49] studied college students, with an average age of 20.5. They reported that users view about two to three surrogates before clicking on a result. Enquiro recruited participants in their 30s and 20s, with some over 40, and reported averages of 3.7 surrogates viewed for young users (under age 35), and an average of 3.8 for users over age 35.

While there are slight differences based on age and education level, both studies consistently showed that individuals rarely visit a second result page. Lorigo et al. [18] reported that users look beyond the top 10 surrogates in less than 5% of the instances. Furthermore, users spend less than two seconds viewing each individual surrogate [49, 48, 51], indicating that the key components of the surrogate are parsed very quickly for a near immediate judgment.

Joachims and Granka also explored how viewing surrogates correlates with the search results that are clicked [54]. The graph in Figure 6 depicts the relationship between the results that are clicked and the surrogates that are viewed. Note the first two results are viewed nearly equally, while the first result is clicked upon disproportionately more often. A later section of this chapter will address this phenomenon in more detail.

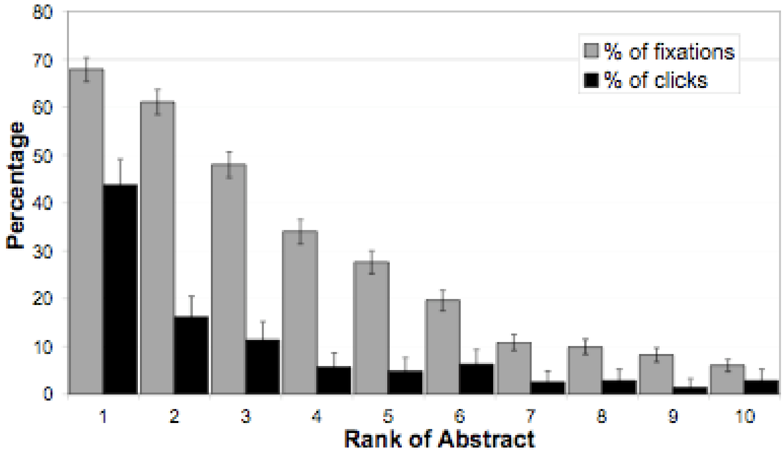

Figure 6. Comparison of the surrogates a user fixates with what a user clicks.

3.5 Viewing Sequence

While we know that in total, only about three to five surrogates are viewed on a search result page, how do users evaluate the options presented to them? More specifically, do users view the surrogates linearly in the order presented, skip around throughout, or backtrack and read a surrogate more than once.

Granka and Joachims et al. [5], [54] both looked at viewing sequence descriptively, recording the order a user viewed surrogates of a particular rank. For each surrogate, this was measured by the fixation at which a user first viewed the surrogate, i.e., at what fixation did a searcher first view the nth-ranked surrogate. Figure 7 indicates that individuals viewed the first and second-ranked results early on, within the second or third fixation. There is a large gap before viewing the third-ranked abstract. Furthermore, the page break also manifests itself in this graph, as the instance of arrival to results seven through ten is much higher than the other six, likely due to the fact that they are displayed below the fold, and few searchers scroll below to read the abstracts.

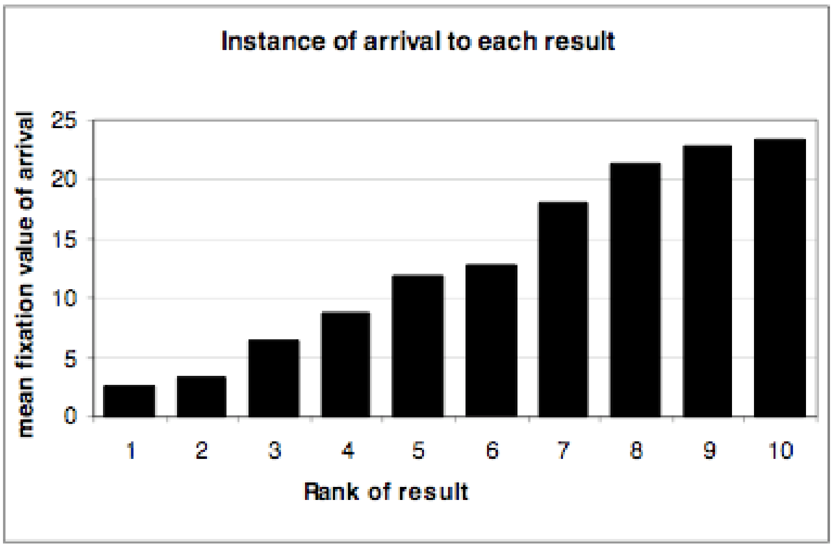

Figure 7. Instance of arrival to individual surrogates, depicting a linear viewing order.

While very useful as a descriptive measure, this analysis does not account for regressions, where a user returns to look at a result a second or third time. It also does not show how users might make pairwise comparisons when selecting a result. There is some existing eyetracking research in the field of consumer marketing which indicates that pairwise comparisons may also be common in selecting search results. Russo and LeClerc [55] outlined several processes that individuals employ before making decisions among various consumer products, the most common strategy being comparisons of adjacent products. It is probable that this same type of behavior also manifests itself during the process of viewing search engine results, where users may feel the need to justify their choice by quickly checking what the next or previous surrogate seems to offer.

Lorigo et al. [18] Extrapolated upon this descriptive approach to viewing order and analyzed the overall sequence of viewing the search results page. Some research was done using scanpath analysis to determine average viewing paths on web stimuli [17, 19]. The work done by [18] was the first to use scanpath analysis to more completely understand viewing on a search result page. They classified viewing behavior on a SERP into one of the following three groups – nonlinear scanning, linear scanning, and strictly linear scanning [18]. Linear scanning may contain jump backs, but adheres to the rule that a surrogate of rank n is not viewed until all surrogates of a smaller rank have also been seen. Strictly linear scanning does not include jump backs or regressions to previously visited surrogates. Also, a scanpath preceding a selection or click is said to be complete if the path contains all surrogates of rank less than or equal to the rank of the selected (clicked) result. Using these scanpaths, we can better characterize how users are likely to make decisions about what results to select.

Lorigo et al. [18] discovered that the query surrogates are viewed in the strict order of their ranking, or strictly linearly, in only about one fifth of the cases, with roughly two thirds of the scanpaths being nonlinear. Hence, even though scanpath sequences are relatively short (visiting 3 surrogates on average), participants were generally not viewing the abstracts in the order intended by the search rank order. Jumps and skips were also short (typically skipping only one surrogate in its path) but prevalent. Visual highlights and cues in the surrogates, such as bolding should not be underestimated as means to grab the attention of the viewer.

Interactions with a SERP occur rapidly. Decisions as to what result to click on or how to refine a query occur in a matter of seconds. While a fixation by fixation analysis may seem tedious, these fixations tell a story about a process as ubiquitous as online search. The example scanpath in Figure 8, mapped to rank values, is: 1 → 2 → 1 → 1 → 2 → 3 → 3 → 2 → 2 → 2, with length of 10 (compare to the reported average length of 16 in [18]). If we ignore repeat fixations within a surrogate, we obtain a compressed sequence of length 6: 1 → 2 → 1 → 2 → 3→ 2. This shows the process of evaluation, with a reported average length of 6 in that same study. If we further ignore repeat surrogates, we obtain the minimal sequence which is the order in which new information was observed: 1 → 2 → 3. Looking at scanpaths in these ways helps to gain a better understanding of the preferences and efficiencies of the participants.

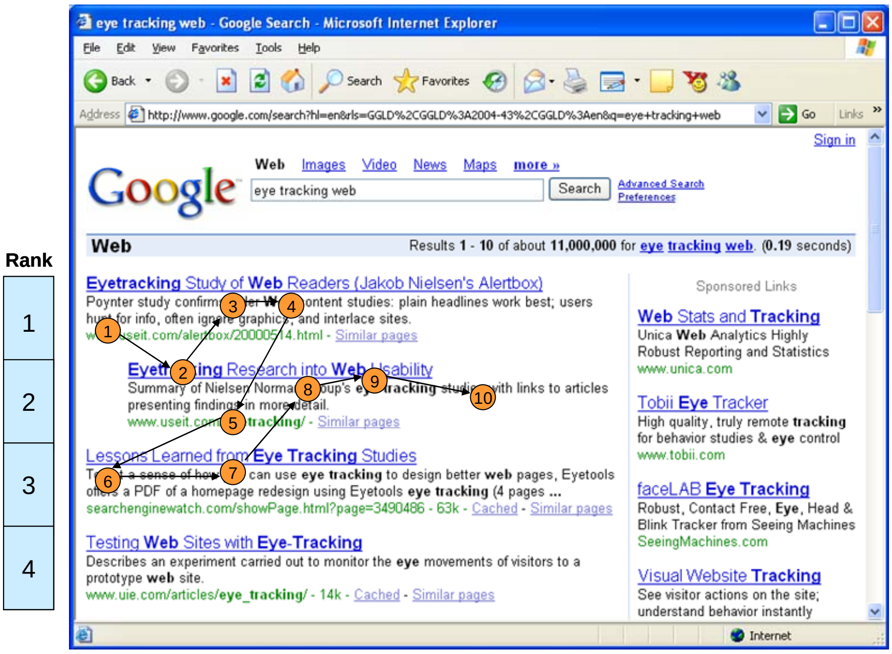

Figure 8. Representation of sample data used for scanpath analysis.

3.6 Result Position

The presentation of the abstracts on the search results page is very likely to influence the manner in which users navigate through and select an online document. In his article, The Intelligent Use of Space, Kirsh [56] outlines several key features about managing and understanding space that may be relevant to the context of document selection in an online information retrieval system. Kirsh points out that one of the goals in organizing both physical and information spaces is to structure the space so that a user's "option set" (the number of viable alternatives available to the user) is reduced and more effectively managed. A reduction in the user's option set is typically accomplished in an information retrieval interface by rank-ordering the retrieved results. By having the system highlight the opportunistic actions that a user should take, the amount of effort that the searcher needs to expend when making a decision is reduced. Because most searchers understand that the top-ranked result has been rated most highly by the information retrieval system, one of the steps to decision-making – organizing the information – is thus eliminated.

All of the studies that have used eyetracking to assess the viewing of search results generally have been able to support the assumption that users rely heavily, almost exclusively, on the ranking offered by the search engine. Joachims took this one step further and switched the order of search results to see how significantly the presentation of results impacts a searcher's viewing behavior [54].

Their study generated three conditions – a normal condition, whereby the search results were left intact as retrieved from the search engine, a reversed condition, where the tenth ranked result was placed in the first position, the ninth in the second, etc, and finally a swapped condition in which only the first and second result were switched. This last condition was included based on findings that the first two results are given nearly equal attention, yet the second result lags disproportionately behind in terms of clickthrough.

Interestingly, through eyetracking analysis, they found that user search behavior was affected by the quality of the results, especially in the completely reversed condition. The swapped condition showed that on average, users were slightly more critical of the results being presented to them, but still clicked on the first result with a greater frequency. In the "reversed" condition subjects scanned significantly more abstracts than in the "normal" condition, and clicked on a lower-ranked result, on average. The average rank of a clicked document in the "normal" condition is 2.66 compared with 4.03 in the "reversed" condition. The study participants did not suspect any manipulation when asked after the session had concluded. The researchers concluded that users have substantial trust in the search engine's ability to estimate the relevance of a page, which influences their clicking behavior [54].

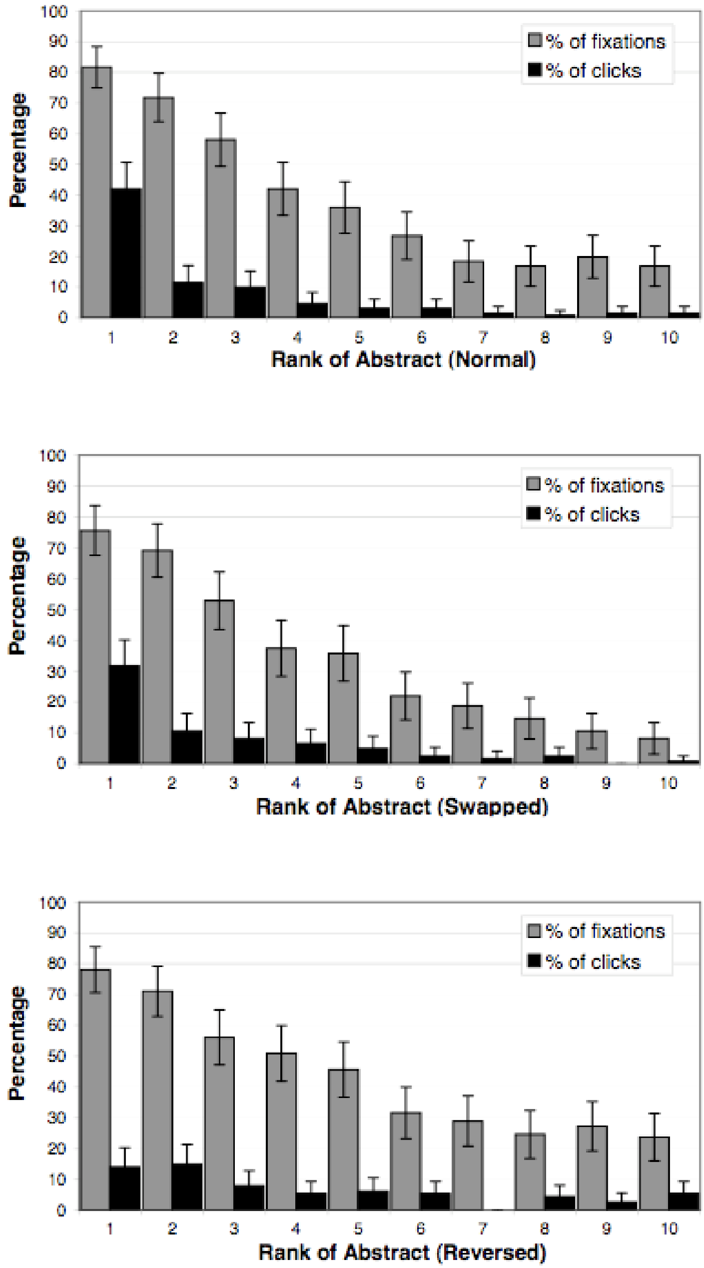

Figure 9. Comparison of what the surrogates a user views and clicks, in the normal, reverse, and swapped condition.

3.7 Task Types and other Influences

Just as the presentation of the surrogates on the SERP page influences navigation behavior, so too can the underlying search task, and also characteristics of each viewer. In his 2002 study of web search, Broder [57] describes three classes of search: navigational, informational, and transactional. In navigational search, the goal is to find a particular web page or URL. In informational search, the goal is to find a particular piece of information that may be on one or multiple web pages. In transactional searches, users are motivated by a desire to conduct a transaction such as making a purchase or trading a stock. Informational search is believed to be the most common type of search performed today [57, 58].

Eyetracking in online search has revealed behavioral differences with respect to task type and user characteristics. Pupil diameter, an indicator of cognitive exertion, was observed to be higher for informational tasks. This is likely because users may need to read more critically when interpreting results for an informational need. This offers behavioral evidence that the cognitive effort involved in informational tasks is greater than that of navigational tasks [49]. Additionally, Lorigo et al. [18] observed informational search tasks took longer than navigational tasks on average, which is not surprising because informational needs are often more complex. However, the time spent on SERPs alone was significantly greater for navigational tasks. This is likely because the answer to a navigational search task is more often found in a surrogate (via the title or URL, indicating the site identity) than for an informational search. The design choice of how much information is placed in a surrogate can have a large impact on user behavior.

We have already noted the difference that age can make on the number of search results viewed. Additionally, there seems to be gender differences in the gaze patterns of the participants, in that males were more linear in their scanpaths than females, and were more likely to view the 7-10th ranked surrogates [18]. These initial findings indicate that the search within the search results may be driven at least in part by user-specific preferences.

Attention to Results Based on Rank. The previous findings have indicated how much attention is given to surrogates on the result page based on their rank. This data is based on all task instances, meaning that if a user did not scroll, their lack of viewing the last half of the page contributes to the lower overall fixation time in those bottom results. However, the figure below offers a different interpretation by only accounting for the instances in which a particular result was looked at. Therefore, if a user did in fact view a surrogate, how much time did she spend in that given surrogate, relative to the others that she viewed.

Figure 10 depicts the amount of time-based attention given to each of the results. There is a dip within the middle-ranked results, indicating that middle results are viewed less exhaustively than the results on either periphery, especially the first two results.

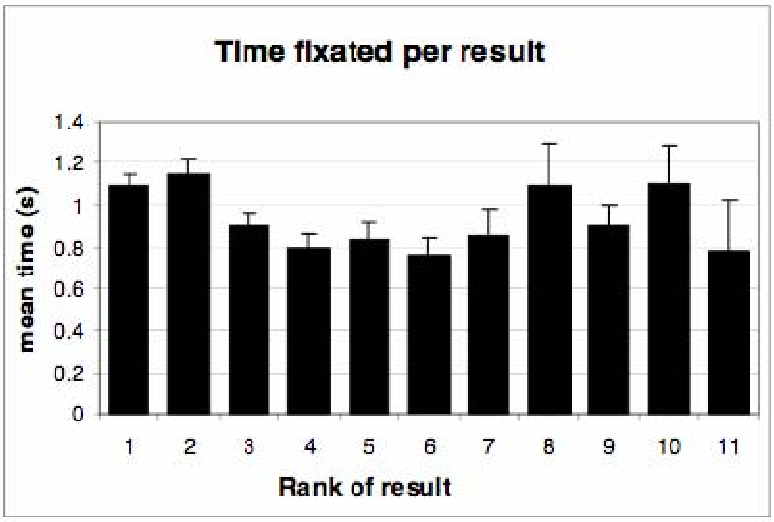

Figure 10. Amount of time spent in each surrogate. Note that the middle results receive lower overall fixation time than the results on the periphery.

In addition to measuring fixation data, knowing how carefully users actually attend to the information in each of the surrogates, in terms of pupil dilation, also tells us whether some results are viewed more attentively, or with more interest, than others. Thus, pupil dilation, as well as total time spent in each surrogate, can be used as measures of interest and cognitive processing.

Figure 11 below depicts the mean pupil dilation on each of the abstracts viewed. This graph follows a trend similar to the one previously described, offering more evidence that middle results are processed less critically. Pupil dilations for the middle ranked surrogates are smaller than the first and last results on SERP, especially the top two.

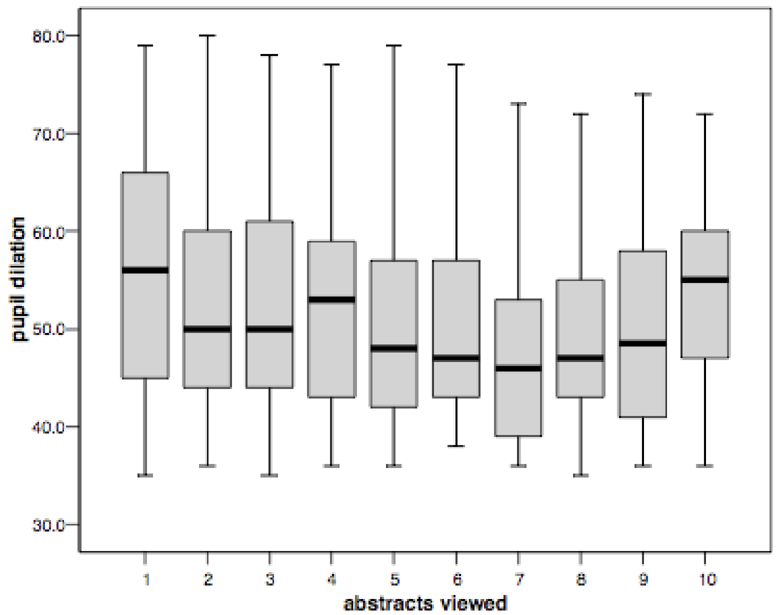

Figure 11. Levels of pupil dilation when viewing surrogates on the SERP. Note that pupil dilation drops after viewing the first ranked surrogate.

3.8 Most Viewed Information in the Surrogate

The majority of search engines format their results with a title, snippet (descriptor of text from the referring web page), and URL. In addition to investigating which of the presented surrogates were viewed, it could also be useful to know what specific content in the surrogates is relied on most to make a relevance judgment. To assess this, attention to each component of the abstract was measured by the number of fixations in each region [49].

Overall, the snippet received the most attention (43% of total fixations), with the title and URL following close behind (30% and 21% of total fixations, respectively). 5% of the total fixations fell on other aspects of the surrogate, including such things as the “cached” link. However, it should be noted that the snippet is proportionately larger than both the URL and title, and is therefore likely to capture more fixations. Furthermore, because the information contained in the snippet is in the form of sentences and full-text content, it is presumably more difficult to cognitively parse than either the titles or URL. The titles and URL can be more readily perceived through an efficient visual scan.

Finally, these findings should be interpreted in light of accuracy limitations inherent to the eyetracker, as previously discussed. Most eyetrackers are only accurate to within a degree of visual angle and thus we cannot be 100% certain of these percentages, but should only use them as a rough estimate for interpreting the impact of each textual component on viewing behavior.

4 Combining Eyetracking with Implicit Measures

Obtaining a low-cost estimate of how users behave in online contexts is desirable, especially for understanding the process of online search. Eyetracking is an expensive analysis tool to use on a regular basis, so in some contexts it would be ideal for researchers to use eyetracking on a periodic basis to enhance their understanding of other low cost, implicit methods. Query logs are an example of one such method, as they capture a great deal of data. In the context of online search, log data can provide a lot of useful information, such as the results a user clicks, how long users spent on a page, etc.

4.1 Eyetracking and Server Log Data

Some information retrieval researchers use server log data to assess the relevance of the documents that are returned to the user. For example, one of the fundamental assumptions is that if a user only clicks through to the third result, then that result must be more relevant than the two listed above it. While assumptions like this (that users look at surrogates in a linear order, users do not look at the documents below the selected one) are the foundation for the algorithms for this work, they had never been behaviorally affirmed. Eyetracking is one way to add value to these algorithms by specifically assessing the likelihood of these assumed behaviors.

Thorsten Joachims was the first to use eyetracking for this very purpose – to more accurately and completely understand the underlying dimensions of server log data [48]. With the help of eyetracking, he demonstrated that there are some problems assuming result clicks can be interpreted as absolute relevance judgments. Instead, they are highly dependent on the user's trust bias towards the ranking of a result ranking. Joachims then offered ways to extract relative feedback by looking at pairwise comparisons between result surrogates.

The only prior work using eyetracking to assess relevance in information retrieval was Salvogarvi, who looked at pupil dilation to determine result relevance, assuming that larger dilations could indicate greater perceptions of relevance. However, this work was strictly within the context of eyetracking, as he did not correlate this method with larger scale data [6].

4.2 Eyetracking and mouse movement data

In addition to combining eye movement data clickthrough data, exploring the correlation between continuous eye and mouse movements could prove to be very useful. Specifically, do the mouse and eye move together around the page? If there is a tight correlation, mouse movements could be a lower cost way of predicting what the eye is observing. Chen et al looked at eye and mouse movements on a set of general web pages. They divided each web page into logical regions, and measured the total time when the mouse and eye were both in a given region. They found that the eye and mouse were only in the same regions slightly above chance, but that if a user made a sudden mouse movement, there was a 84% chance that the user also looked to where the mouse moved.

Rodden and Fu [59] explored this relationship of mouse to eye movements within the context of Google search result pages, and found that there is indeed some relationship between the two measures, but likely not enough to use mouse movements as a substitute for eyetracking. They used scanpath analysis, similar to Lorigo et al. [18] did, but extended their analysis to compare mouse paths with the eye scanpath. They noted some interesting relationships between the mouse and eye, classifying three types of mouse and eye correlations. First, searchers sometimes use the mouse to mark an interesting result, how users may keep the mouse still while reading, or ultimately, move the mouse along as they read text.

5. Conclusion

Due to technological advances, eyetracking is now a viable option for many user experience researchers and industry practitioners. Eyetracking has been used to understand user behaviors in a number of online contexts, including news, homepages, and search results. While some challenges remain that limit ubiquitous use of eyetracking (namely the high cost and lack of efficient analysis tools), significant contributions have already been made which would not have been possible without this tool.

Specifically in online search, researchers now have more a more complete interpretation of what happens *before* a user selects one of the search results presented to them. Prior to eyetracking, researchers could only rely on server log data to understand distinct user actions. With eyetracking, researchers can now get a deeper cognitive and behavioral understanding of how individuals process the information presented to them online. This will add a valuable level of understanding, particularly with the growth in individuals relying on the internet for information.

Acknowledgments. Many thanks to Geri Gay, Helene Hembrooke, Thorsten Joachims, Fabio Pellacini, Kerry Rodden, and Daniel Russell for their invaluable contributions and discussions.

References

1. Stanford Poynter Project. Front Page Entry Points. http://www.poynterextra.org/et/i.htm

2. Benway, J. and Lane, D. Banner blindness: web searchers often miss "obvious" links. Internetworking, 1.3. URL

3. Dreze, X. and Hussherr, F.: Internet Advertising: Is Anybody Watching? Journal of Interactive Marketing, 17, 4 (2003) 8-23

4. Goldberg, J.H., Stimson, M.J., Lewenstein, M., Scott, N., & Wichansky, A.N. (2002). Eye tracking in web search tasks: Design Implications.

5. Granka, L., Joachims, T., Gay, G.: Eye-Tracking Analysis of User Behavior in WWW-Search, Poster Abstract, Proceedings of the Conference on Research and Development in Information Retrieval (SIGIR) (2004)

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

7. Tobii Technology: http://www.tobii.com/products_-_services/testing_and_research_products/tobii_x50_eye_tracker/

8. Applied Science Laboratories (ASL): http://www.a-s-l.com/

9. Rayner, K.: Eye movements in reading and information processing: Twenty years of research. Psychological Bulletin 124 (1998) 372-422

10. McConkie, G., Loschky, L.: Perception onset time during fixations in free viewing. Behavior Research Methods, Instruments, & Computers 34 4 (2002) 481-490.

11. Salvucci, D. D., & Goldberg, J. H.: Identifying fixations and saccades in eye-tracking protocols. In Proceedings of the Eye Tracking Research and Applications Symposium (pp. 71-78). New York: ACM Press (2000).

12. Just, M.A. & Carpenter, P.A.: A theory of reading: From eye fixations to comprehension. Psychological Review, 87 (1980) 329-354

13. Viviani, P. Chapter 8. In Kowler, E. (Ed.) Eye Movements and Their Role in Visual and Cognitive Processes. Amsterdam: Elsevier Science (1990)

14. Duchowski, A. T. A Breadth-First Survey of Eye Tracking Applications. Behavior Research Methods, Instruments, & Computers (BRMIC), 34(4), November 2002, pp.455-470.

15. Hess, E. and Polt, J.: Pupil size in relation to mental activity during simple problem-solving. Science, 143 (1964) 1190-1192.

16. Hess, E. and Polt, J.: Pupil size as related to interest value of visual stimuli. Science, 132, 3423 (1960) 349-350.

17. Josephson, S., and Holmes, M.E.: Visual Attention to Repeated Internet Images: Testing the Scanpath Theory on the World Wide Web. Proceedings of Eye Tracking Research & Applications: Symposium ACM SIGCHI (2002) 43-51

18. Lorigo, L., Pan, B., Hembrooke, H., Joachims, T., Granka, L., Gay, G.: The Influence of Task & Gender on Search & Evaluation Behavior Using Google. Information Processing & Management (2006)

19. Hembrooke, H., Feusner, M., & Gay, G.: Averaging Scan Patterns and What They Can Tell Us. ETRA (2006)

20. Yarbus, A.L. 1967. Eye Movements and Vision. Plenum Press: New Yorkvb

21. Pellacini, F., Lorigo, L., & Gay, G. (2006). Visualizing Paths in Context, Technical Report #TR2006-580, Dept. of Computer Science, Dartmouth College.

22. Leigh, R.J. and Zee, D.S., 1991. The Neurology of Eye Movements, 2nd ed F.A. Davis Co, Philadelphia.

23. Goldberg, J., Kotval, X.: Computer interface evaluation using eye movements: methods and constructs. International Journal of Industrial Ergonomics 24 (1999) 631-645

24. Wooding, D.: Eye movements of large populations: II. Deriving regions of interest, coverage, and similarity using fixation maps. Behavior Research Methods, Instruments, & Computers 34 (4) (2002) 518-52

25. Santella, A., DeCarlo, D.: Robust Clustering of Eye Movement Recordings for Quantification of Visual Interest. Proc. of Eye Tracking Research & Applications (2004) 27-34

26. Granka, L., Rodden, K. Incorporating Eyetracking into User Studies at Google. Workshop Position paper presented at CHI 2006, ACM.

27. McCulloch, C. and Searle, S.: Generalized, Linear, and Mixed Models. John Wiley and Sons, New York. (2001)

28. Levenshtein, V.: Binary codes capable of correcting deletions, insertions, and reversals. Doklady Akademii Nauk SSSR, 163(4):845-848, 1965 (Russian). English translation in Soviet Physics Doklady, 10(8):707-710, 1966.

29. Needleman, S., Wunsch C.: A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol Mar;48(3) (1970) 443-53

30. Pan, B., Hembrooke, H., Gay, G., Granka, L., Feusner, M., & Newman, J.: The Determinants of Web Page Viewing Behavior: An Eye-Tracking Study. Proc. of Eye Tracking Research & Applications (2004) 147-154

31. West, J., Haake, A., Rozanski, E., Karn, K.: eyePatterns: Software for Identifying Patterns and Similarities Across Fixation Sequences. Proc. of Eye Tracking Research & Applications (2006) 149-154

32. Hacisalihzade, S., Stark, L., Allen, J.: Visual Perception and Sequences of Eye Movement Fixations: A Stochastic Modeling Approach. IEEE Transactions on Systems, Man, and Cybernetics 22 3 (1992) 474-481

33. Ponsoda, V., Scott, D., Findlay, J.M.: A probability vector and transition matrix analysis of eye movements during visual search. Acta Psychologica 88 (1995) 167-185

34. Salvucci, D.: Inferring Intent in Eye-Based Interfaces: Tracing Eye Movements with Process Models. CHI (1999) 254-261

35. Pieters, R., Rosbergen, E., Wedel, M.: Visual Attention to Repeated Print Advertising: A Test of Scanpath Theory. Journal of Marketing Research 36 (1999) 424-438

36. OSHA www.osha.gov/SLTC/etools/computerworkstations/components_monitors.html

37. Uwano, H., Nakamura, M., Monden, A., Matsumoto, K.: Analyzing Individual Performance of Source Code Review Using Reviewer' Eye Movement. Proc. of Eye Tracking Research & Applications (2006) 133-140

38. Tobii Technology, Clearview Software, http://www.tobii.com/products_-_services/testing_and_research_products/tobii_clearview_analysis_software/

39. Lankford, C.: Gazetracker: Software Designed to Facilitate Eye Movement Analysis. Proceedings of the Eye Tracking Research & Applications Symposium: ACM SIGCHI (2000) 43-51

40. SMI http://www.smi.de/iv/

41. Bates, R. and Spakov, O.: D2.3 Implementation of COGAIN Gaze Tracking Standards. Communication by Gaze Interaction (COGAIN), IST-2003-511598: Deliverable 2.3 (2006) Available at http://www.cogain.org/results/reports/COGAIN-D2.3.pdf

42. Holmes, M. E.: WinPhaser User's Manual: Version (Version 1.0c) [computer software] (1996)

43. Joachims, T.: Optimizing search engines using clickthrough data. Proceedings of the ACM Conference on Knowledge Discovery and Data Mining, ACM (2002) 132-142

44. Drucker, H., Shahrary, B., and Gibbon, D.: Support vector machines: relevance feedback and information retrieval. Information Processing and Management 38 3 (2002) 305-451

45. Spink, A.: A user-centered approach to evaluating human interaction with Web search engines: an exploratory study. Information Processing and Management 38 (2002) 401-426

46. Hembrooke, H., Granka, L., & Gay, G. The Effects of Expertise and Feedback on Query Formation and Expansion: Developing a Typology of Search Term Strategy Use. Journal of the American Society for Information Science. (2005)

47. Hsieh-Yee, I.: Research on Web Search Behavior. Library and Information Science Research 23 (2001) 167-185

48. Joachims, T., Granka, L., Pan, B., Hembrooke, H., Radlinski, F., Gay, G. Evaluating the Accuracy of Implicit Feedback from Clicks and Query Reformulations in Web Search. Transactions on Information Systems (In press)

49. Granka, L.: Eye-R: Eye-Tracking analysis of user Behavior in online Search. Masters Thesis, Cornell University Library Press (2004)

50. Nielsen, J.: F-Shaped Pattern for Reading Web Content. Altertbox April 17, (2006)

51. Enquiro: Enquiro Eye Tracking Report I: Google, July 2005, http://www.enquiro.com/research.asp

52. Rodden, K. & Fu, X.: Eyetracking Analysis of Google Web Search. Internal Research, Google Inc. 2006

53. Aula, A. Majaranta, P. & Raiha, K-K.: Eye-tracking reveals the personal styles for search result evaluation. In M.F. Constabile & F. Paterno (Eds.) Proceedings of the IFIP TC 13 International Conference on Human-Computer Interaction (INTERACT), Springer-Verlag, (2005) 1058-1061

54. Joachims, T., Granka, L., Pang, B., Hembrooke, H., & Gay, G.: Accurately Interpreting Clickthrough Data as Implicit Feedback. Proceedings of the Conference on Research and Development in Information Retrieval (SIGIR) (2005)

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

56. Kirsh, D. The Intelligent use of space. Artificial Intelligence, 73, (1-2), (1995) 31-68

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

58. Rose, D.E., & Levinson, D.: Understanding user goals in web search, Proceedings of the Thirteenth Annual World Wide Web Conference. New York: ACM (2004) 13-19

59. Rodden & Fu. X: Exploring how mouse movements relate to eye movements on web search results pages. Proceedings of the Conference on Research and Development in Information Retrieval (SIGIR) (2007) (Submitted)