User See, User Point:
Gaze and Cursor Alignment in Web Search

Jeff Huang
University of Washington
Seattle, WA 98195
chi@jeffhuang.com

Ryen W. White
Microsoft Research
Redmond, WA 98052
ryenw@microsoft.com

Georg Buscher
Microsoft Bing
Bellevue, WA 98004
georgbu@microsoft.com

ABSTRACT

Past studies of user behavior in Web search have correlated eye-gaze and mouse cursor positions, and other lines of research have found cursor interactions to be useful in determining user intent and relevant parts of Web pages. However, cursor interactions are not all the same; different types of cursor behavior patterns exist, such as reading, hesitating, scrolling and clicking, each of which has a different meaning. We conduct a search study with 36 subjects and 32 search tasks to determine *when* gaze and cursor are aligned, and thus *when* the cursor position is a good proxy for gaze position. We study the effect of time, behavior patterns, user, and search task on the gaze-cursor alignment, findings which lead us to question the maxim that "gaze is well approximated by cursor." These lessons inform an experiment in which we predict the gaze position with better accuracy than simply using the cursor position, improving the state-of-the-art technique for approximating visual attention with the cursor. Our new technique can help make better use of large-scale cursor data in identifying how users examine Web search pages.

Author Keywords

gaze; cursor; eye-tracking; search examination behavior

ACM Classification Keywords

H.5.4 Information Interfaces and Presentation: Hypertext / Hypermedia—*Navigation*

General Terms

Human Factors, Measurement, Experimentation

INTRODUCTION

Billions of people navigate the Web as part of their daily lives by looking, finding, reading, pointing, and clicking. Because the flow of using a site is such a fundamental experience for so many people, usability professionals and site designers seek to optimize the experience by analyzing which parts of the page grab a visitor's attention, and what information people read on the page. To achieve this, they conduct laboratory studies using eye-tracking equipment to track users' gaze while they navigate the site. Research studies have shown

gaze to be useful in determining what people are reading from their fixation [5, 26], in determining salient regions of Web pages [4], and in identifying the effects of changes to search engine results pages (SERPs) [13]. But could the mouse cursor be a cheap and scalable alternative to eye-tracking for the same purposes? If so, cursor tracking systems can be deployed on any website without the need for an eye-tracker.

Most people use a mouse cursor to interact with Web pages; this interaction data can be efficiently collected at scale in a natural setting without disrupting the user [2, 20, 22]. There are reports that the cursor is a suitable substitute for eye-tracking for determining people's attention and examination behavior. Academic studies [1, 2, 31] and commercial offerings (e.g., ClickTale [clicktale.com], Lucky Orange [lucky-orange.com], Observe.It [observe.it], and Mouse Eye Tracking [met.picnet.com.au]) have analyzed the cursor in usability settings to learn about engagement on Web pages. Past research has found a correlation between gaze and cursor positions [9, 12, 19, 22, 27, 28] and that cursor movements can be useful for determining relevant parts of the Web page with varying degrees of success [15, 21, 22, 29]. However, cursor interaction spans a variety of behaviors [1, 12, 25, 27, 28] including reading, hesitating, highlighting, marking, and actions such as scrolling and clicking. Identifying these behaviors is a prerequisite to understanding what meaning the cursor interactions convey.

In this work, we aim to further quantify the relationship between cursor and gaze. Prior work has shown that gaze and cursor are correlated but the goal of our work is to determine *when* gaze and cursor are aligned. We want to know when the cursor position is a good proxy for gaze position and the effect of various factors such as time, user, cursor behavior patterns, and search task on the gaze-cursor alignment. Through this, our research will contribute to knowledge of how people use their cursor and how they examine Web search pages. The lessons from this analysis also inform an experiment in which we *predict* the gaze position with better accuracy than simply using the corresponding cursor position. In addition to the cursor position, we also use dwell time[1], temporal features relating to cursor behavior, and future cursor positions to determine the current gaze position.

The findings from this work can be applied to many practical situations. Predicting gaze will improve state-of-the-art techniques for approximating visual attention with the mouse cursor. Current commercial cursor-tracking systems merely

---

[1]Dwell time here is the page dwell time as typically measured in information retrieval literature, rather than cursor dwell time.

present heatmaps of the cursor movements and claim that the user attention is proximal to the hot areas of the map. Our findings will inform these and other applications that remotely collect large amounts of cursor data at scale by showing how to more effectively use this data. More broadly, our work supports collecting user attention data in situations where eye-tracking equipment is costly, such as for large numbers of Web users. One particular domain that may benefit from this research is Web search, in which searcher examination behavior represented by probabilistic models [8, 16] may be improved by uncovering the latent variables representing which search results the users examine.

We study the gaze-cursor alignment of 36 subjects and 32 search tasks with an eye-tracker in a controlled lab setting. The findings show that while subjects vary their gaze-cursor alignment substantially, they lag their cursor behind their gaze by at least 250 ms and on average by 700 ms. Both dwell time and the user's personal style affect the distance between the cursor and gaze positions. This distance is longer when the cursor is inactive, shorter when the cursor is used to help examine or read the page, and even shorter when the user is performing an action. Finally, we predict the gaze position using cursor features and achieve 23.5% more accuracy than simply using the corresponding cursor position alone.

RELATED WORK

Two lines of research relate to this work. One focuses on inferring user interest and intentions directly from the user's interactions with a SERP or a Web page, specifically based on cursor movements, clicks, and gaze. The second explores the relationship between cursor position and gaze position in order to infer users' visual attention from cursor movements.

Implicit Interest Indicators

In early work, Goecks and Shavlik modified a Web browser to record themselves browsing hundreds of Web pages [15]. They found that a neural network could predict variables such as the amount of cursor activity on the SERP, which they considered surrogate measurements of user interest. Claypool et al. [10] developed the "curious browser", a custom Web browser that recorded activity from 75 students browsing over 2,500 Web pages. They found that cursor travel time was a positive indicator of a Web page's relevance, but could only differentiate highly irrelevant Web pages. They also found that the number of mouse clicks on a page did not correlate with its relevance, despite the intuition that clicks represent links that users found appealing. Hijikata [21] used client-side logging to monitor five subjects browsing a total of 120 Web pages. They recorded actions such as text tracing and link pointing using the cursor. The findings showed that these behaviors were good indicators for interesting regions of the Web page, around 1.5 times more effective than rudimentary term matching between the query and regions of the page. Shapira et al. [29] developed a special Web browser and recorded cursor activity from a small number of company employees browsing the Web. They found that the ratio of mouse movement to reading time was a better indicator of page quality than cursor travel distance and overall length of time that users spent on a page.

In the search domain, Guo and Agichtein [17] captured mouse movements using a modified browser toolbar and found differences in cursor travel distances between informational and navigational queries. Furthermore, a decision tree could classify the query type using cursor movements more accurately than using clicks. Guo and Agichtein also used interactions such as cursor movement, hovers, and scrolling to accurately infer search intent and interest in search results [18]. They focused on automatically identifying a searcher's research or purchase intent based on features of the interaction. Buscher et al. [7] also used scrolling to infer user interest and compared it to gaze tracking feedback. They found that scrolling behavior in connection with information on the browser's viewport could be as effective as gaze tracking feedback in search scenarios comprising query expansion. By applying a reading detection method, Buscher et al. [5] used gaze tracking features directly to infer user interest and show that this can yield great improvements when personalizing search. Buscher et al. also demonstrated the value of gaze information to build models for predicting salient regions of Web pages [4]. Finally, Cole et al. [11] built reading models operating on gaze tracking data to investigate information acquisition strategies of the searcher for different search tasks.

Gaze-Cursor Relationship

Another line of research examines the relationship between eye gaze and cursor positions. An early study by Chen et al. [9] measures this relationship by recording 100 gaze and cursor positions from five subjects browsing the Web. They showed that the distance between gaze and cursor was markedly smaller in regions of encountered pages to which users attended. Liu and Chung [25] recorded cursor activity from 28 students browsing the Web and noticed patterns of viewing behaviors, including reading by tracing text with the cursor. Their algorithms were capable of predicting users' cursor behaviors with 79% accuracy. Hauger et al. conducted a study using a similar method to ours, but over instructional pages about the game of "Go" [20]. They found that gaze and cursor positions were better correlated when the cursor was in motion and in sessions comprising a higher proportion of motion. Hauger et al. were able to predict which paragraph a user was reading, and to what extent, with 79% accuracy.

Other work has focused on the relationship between cursor and gaze in search tasks. In a study involving 32 subjects performing 16 search tasks each [27, 28], Rodden et al. identified a strong alignment between cursor and gaze positions. They found that the distance between cursor and gaze positions was larger along the x-axis than the y-axis, and was generally shorter when the cursor was placed over the search results. Rodden et al. also observed four types of mouse cursor behaviors: neglecting the cursor while reading, using the cursor as a reading aid to follow text (either horizontally or vertically), and using the cursor to mark interesting results. Guo and Agichtein [19] reported similar findings in a smaller study with ten subjects performing 20 search tasks each. Like Rodden et al., Guo and Agichtein observed that distances along the x-axis tended to be larger than the distances along the y-axis. They could predict with 77% accuracy when gaze and cursor were strongly aligned using cursor features.

The research presented in this paper extends previous work in several ways. The focus of previous work was in examining the relationship of cursor and gaze position by measuring the distance between them. The conclusions were that a strong correlation existed between them, however there was little further differentiation of alignment along a temporal dimension and no differentiation by cursor behavior patterns. Here we look at the *degree of alignment* at different points in time, quantitatively analyzing behaviors made using the cursor, and we are *predicting* the gaze position using a variety of cursor behavior and time features. In addition, our analysis of the gaze-cursor relationship involves a larger number search tasks and subjects than prior studies, giving us more leverage in feature comparisons.

METHOD

We used a Tobii x50 eye tracker with 50 Hz tracking frequency and an accuracy of $0.5°$ visual angle (corresponding to 16 pixels in our setting) on a 1280×1024 resolution 17 inch monitor (96dpi) and 1040×996 resolution Web browser. Cursor and gaze coordinates were collected in an eye-tracking study with 38 subjects (21 female, 17 male), recruited from a user study pool. They ranged in age between 26 and 60 years ($M = 45.5$, $SD = 8.2$), and possessed a wide variety of backgrounds and professions. Two subjects had incomplete data due to technical issues and were dropped from the analysis.

We prepared a set of 32 Web search tasks (information needs) which each subject completed on the Bing search engine in a randomized order. Half of the tasks were navigational (i.e., they had to find a specific Web page) and half were informational (i.e., they had to find factual information). Each task started with a description of what subjects should look for on the Web. They had to start searching with a predefined query for each task (e.g., task = "What are some side-effects of Ibuprofen?", predefined query = "ibuprofen side effects") that we generated from the task, but were then free to interact with the search results, browse the Web and search further. The browser cache and cookies were cleared after each subject to prevent subjects from noticing previously viewed pages (from hyperlinks turning purple) and search engine personalization effects. Going through an eye tracker calibration phase in the beginning, completing all 32 search tasks, as well as filling in a demographics questionnaire in the end took about one hour per subject.

Gaze and cursor positions were recorded for each SERP as well as subsequent Web pages (i.e., pages visited after clicking on a search result). In total, we collected data for 1,210 search tasks, which included 1,336,647 gaze positions, and 87,227 cursor positions (whenever the subject moved the cursor). Additional details about the experimental procedure are described in Buscher et al. [6]. Gaze-specific findings on this data set, unrelated to cursor features, have been reported elsewhere [6, 14]. Huang et al. [22] used this data to measure gaze-cursor alignment (Figure 1), showing that there is some correlation between gaze and cursor positions, but with substantial variation. Other studies measuring gaze-cursor alignment [19, 27] have presented evidence of alignment with charts nearly identical to Figure 1.
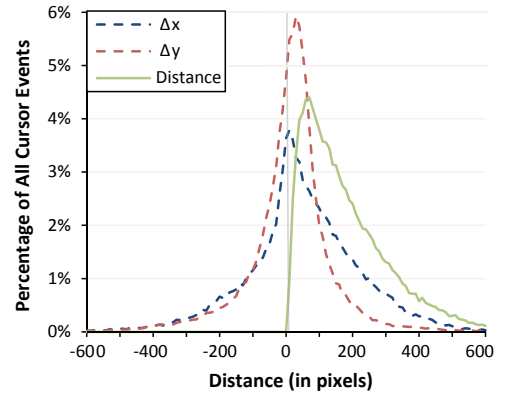

Figure 1. $\triangle x$, $\triangle y$, and Euclidean distance plotted in a frequency distribution for the search results page.

In the search logs, the gaze positions were recorded approximately every 20 ms, whereas cursor positions were recorded approximately every 100 ms. Gaze positions are estimated by the eye-tracker and given the saccades (rapid movements) of the eye, recording at a higher frequency can give more accurate positions; but the cursor position is an exact value and recording at 10 Hz is sufficient—the cursor is not radically changing directions at sub-second speeds such that we cannot interpolate them. We had tested different frequencies of interpolation and found negligible differences.

Since cursor and gaze events did not necessarily have identical timestamps, a gaze position was interpolated for every cursor position. Interpolation was performed by computing gaze x and y coordinates weighted by the coordinates of the nearest gaze coordinates before and after the cursor position. For example, the interpolated x-coordinate for eye gaze is computed as,

$$x_i = x_0 + (x_1 - x_0)\frac{t_i - t_0}{t_1 - t_0} \qquad (1)$$

where t_i is the time for the corresponding cursor position, x_0 is the gaze's x-coordinate preceding the cursor position, recorded at time t_0, and x_1 is the gaze's x-coordinate following the cursor position, recorded at time t_1. The interpolated y-coordinate was computed the same way, substituting x for y in the above equation. To reduce interpolation inaccuracies due to noise from the eye-tracker, cursor positions were only captured if they occurred between gaze positions that were at most 100 ms apart.

ALIGNMENT EFFECTS FROM TIME, USER, AND TASK

To begin our analysis, we view gaze-cursor alignment through the lens of time, user, and search task. Different people behave differently on the Web in the queries they issue [33], how they gaze at the page [3, 14], and how they interact with the page [19]. The amount of time spent reviewing the search results page can affect where they are pointing and where they are looking. Additionally, we investigate the possibility that gaze and cursor may not be most aligned at each instance in time, but may be better aligned when one variable is temporally shifted.
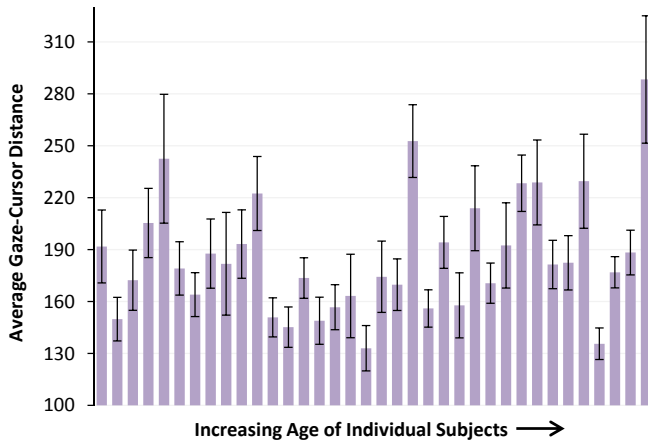
Figure 2. The average gaze-cursor distance for each subject with error bars representing the standard error. The variance between subjects is high ($SD = 33.9$). Subjects are sorted by ascending age so that oldest subjects are on the right.
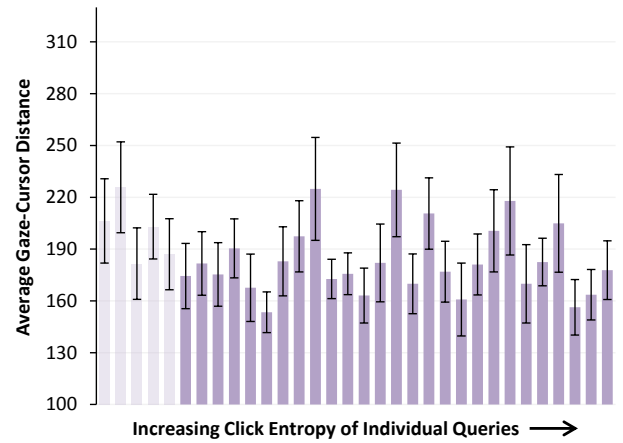

Figure 3. The average gaze-cursor distance for each search task with error bars representing the standard error. The variance between search tasks is modest ($SD = 20.2$). Queries are sorted by ascending click entropy so queries with more diverse result clicks are on the right; five queries with unknown entropy are on the left (shaded lighter).

User and Task Effects

People visually examine Web pages differently [4], and have individual styles in their control of the cursor. Past findings have shown that users behave differently depending on search tasks (e.g., between informational and navigational task types [13]). Guo and Agichtein analyzed gaze-cursor alignment [19] on a smaller subject pool and found differences in user (up to twice the average distance between gaze and cursor for some users) and search task type (navigational vs. informational). Here we explore whether search task or individual differences (including age and gender) have a stronger effect on alignment. Larger variances in alignment would indicate that a search system might have difficulty predicting alignment for users or queries that have not occurred before.

To study alignment differences among subjects, we macro-averaged the gaze-cursor distance across their queries (i.e., took the average of the average for each of their queries, thus giving equal weight to each query). The results in Figure 2 show that subjects are fairly distinct in terms of gaze-cursor alignment. Some keep their cursor within about 130 px of their gaze, while others average about 280 px. The standard deviation representing the variation among subjects is $SD = 33.9$. We checked for gender differences in gaze-cursor distance using a two-tailed t-test and found no statistically significant effect ($t(34) = 1.31, p = 0.20$). There was a Spearman correlation of $\rho = 0.22$ between age and gaze-cursor alignment, but this was also not statistically significant ($N = 36, p = 0.18$). Therefore, we conclude that whether the subject tracks their gaze closely with their cursor is more likely to stem from personal habits rather than age or gender.

Next we look at the average differences in gaze-cursor alignment for different search tasks. Since the subjects were given predefined queries to begin with, we filtered out reformulated queries since they may be reflective of personal style or individual search skills. A search task is therefore represented by a single query in this analysis. For each search task, we first macro-averaged the gaze-cursor distance across subjects to normalize the data, then took the mean of this value as the average distance for each task. The search task averages are shown in Figure 3, which reveals that there are modest differences between tasks. The alignment distance ranges from about 150 px to about 220 px, and the standard deviation among search tasks is $SD = 20.2$.

To gain some insight into the attributes of the search tasks, we computed the click entropy for each query, a measure of result click diversity [33]. The click entropy is computed as the Shannon entropy of the click distribution on the search result links. This was done to look for correlation between click entropy and gaze-cursor alignment; we thought it was possible that gaze and cursor would align differently depending on query type. (Typically, queries with high click entropy are informational or more ambiguous, and navigational otherwise [30].) We used the past year's search logs of the Bing search engine to find the click distributions. Five queries had not appeared in the search logs in the past year and thus had unknown entropy. From the remaining queries, we found no Spearman correlation between click entropy and gaze-cursor alignment ($\rho = 0.01, N = 27, p = 0.96$). This contrasts findings from Guo and Agichtein's study [19] which found alignment differences between navigational and informational queries.

The standard deviation in gaze-cursor alignment across means for different subjects is higher than across means for different search tasks. A Levene's test for homogeneity of variance shows that the differences in variance are statistically significant (Levene statistic[2] $= 4.529, p = 0.037$). This suggests that users have individual preferences and that these differences are stronger than the differences between search tasks. We experimented with not normalizing the data when computing the mean averages (i.e., micro-averaging), but the differences in standard deviation between search task and subject were similar.

---

[2]Levene's test is typically used to compare the variance between two sources rather than between two sets of macro-averaged values.
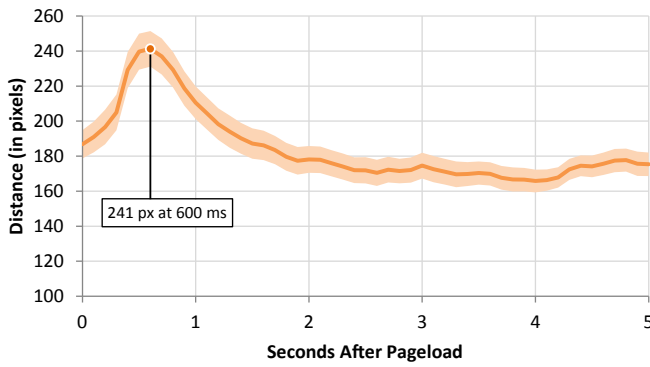
Figure 4. The average gaze-cursor distance at 100 ms intervals after the SERP loads, macro-averaged over subjects. The distance is low just as the page loads, increases at 0.5–1 seconds, then decreases. The shaded area is the region representing the standard error of the mean.

Temporal Effects

Users examine the SERP in a sequence of evolving behaviors. A user may possess the habit of quickly scanning the page first to see what kinds of items are on the page, then skimming it quickly to see if there is an answer to their information need while neglecting the cursor, then finally reading the text word-by-word. As the time spent dwelling on the SERP increases, the alignment between gaze and cursor may change due to the dynamism of the behaviors.

Averaging gaze and cursor distances for different dwell times, Figure 4 shows the relationship between gaze and cursor over time for the first five seconds following the pageload. Indeed, the time since the page has loaded has an effect on alignment. Specifically, the alignment distance ranges from 170 px to almost 240 px. Right when the page loads, gaze and cursor are closely aligned, perhaps from the previous action that led to the page. The peak at 240 px is within one second of the page loading, suggesting that the subject may first scan the displayed page without moving their cursor. The gaze-cursor alignment narrows after about two seconds when the subject may start to examine the page more closely and perhaps prepare to click a link. While the subject's actions at this stage change from query-to-query, the aggregate alignments provide clues of typical examination behavior.

This behavior led us to ask—given that we see the eye moving within a second of the page loading, does the cursor move as quickly? We arrived at three credible hypotheses: a) the user uses the cursor as a reading aid [28] and so the eye follows the cursor, b) the gaze and cursor positions are best aligned at each given point in time, c) or the cursor follows gaze because the user looks at something and then moves their cursor to interact with it. The last hypothesis is consistent with findings from early work by Ware and Mikaelian [32] that showed that the eye fixation was a faster method for target selection than the mouse. While target selection is not the objective of this study, the difference between eye and mouse speed may cause the cursor to lag behind the gaze.

We used a technique similar to cross-correlation in signal processing that computes when two waveforms are most aligned [23]. We first interpolated cursor and gaze positions at 50
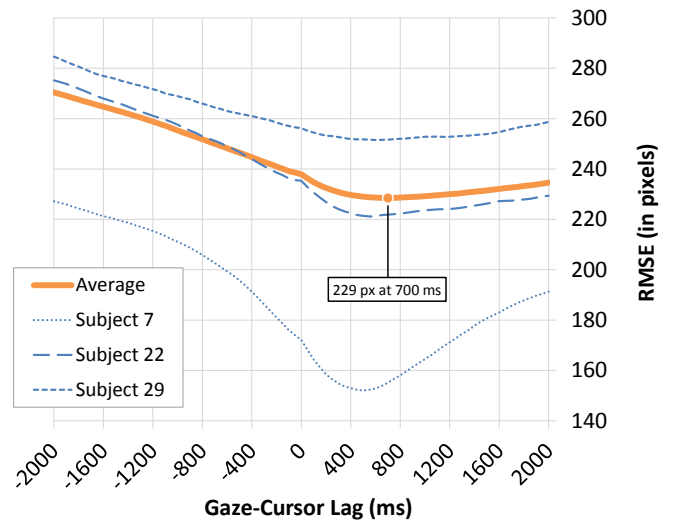

Figure 5. The root-mean-square error for gaze and cursor distance at different intervals (in 50 ms increments) of gaze-cursor lag, representing how well gaze positions correlate with future and past cursor positions. The thick solid line plots the RMSE macro-averaged over subjects, and thin dashed lines plot the RMSE for three example subjects.

ms intervals again using Equation 1. This equalized the periods between positional data points, computing the cursor positions at different time shifts (e.g., 50 ms, 100 ms, 150 ms, ...) after each gaze position. Then we computed the root mean square error (RMSE) between gaze and cursor[3] for each shifted time interval for each subject. The cursor and gaze positions are considered most correlated at the time shift with the lowest RMSE.

Figure 5 shows the macro-averaged RMSE values across subjects compared with three example subjects. The first thing to note is that for the macro-averaged values, shifting the cursor positions 700 ms into the future minimizes the RMSE. This means mouse cursor lags behind eye gaze by about 700 ms, so the user looks at something and then (almost one second) later their cursor is moved to that location on the page. Temporal alignments vary depending on the user and query session; the per-subject RMSE values showed that different people had different delays in moving their cursor to their gaze. In the same manner, the cursor lagged behind the gaze for each individual subject; the inverse situation—gaze lagging behind the cursor—did not occur, refuting the hypothesis that some people lead with the cursor when examining the page. Still, some subjects were quick, moving their cursor to their gaze in 350 ms (one subject had a minimum RMSE at 250 ms), while others took over one second.

CURSOR BEHAVIOR PATTERNS

Several studies have reported different behavioral patterns when using the cursor. Rodden et al. observed several types of cursor behavior: neglecting the cursor while reading, using the cursor as a reading aid to follow text, and using the cursor to mark interesting results [28]. Liu and Chung also noticed

---

[3]Note that RMSE differs from the average alignment distance metric we use throughout the paper; RMSE takes the root of the mean of the squared distances which is not the mean distance.

patterns of cursor behaviors, including reading by tracing text [25]. Arroyo et al. observed users hesitating and reading with the cursor [1]. Mueller and Lockerd found users hesitating before clicking and resting the cursor on white space [12]. Finally, Claypool et al. observed the following cursor behaviors: ignoring the cursor, examining the page using the cursor, following the text with the cursor, and using it to interact with the page or browser [10]. Our taxonomy of cursor behavior resembles that reported by Claypool et al.

Each of the previously mentioned studies discusses the cursor behavior qualitatively and most do not have corresponding gaze data. Our study involved more subjects and search tasks than the largest prior study, facilitating a quantitative analysis of the cursor behaviors. Informed by prior work and our own qualitative observations of user interactions with the SERP, we separated cursor behaviors into four categories:

Inactive cursors are not moving and are ignored by the user for some time.

Reading cursors are used to follow the text while the user is reading the page.

Action cursors are used when the user is about to perform an action (click on a link, edit the query in the search box, drag the scrollbar, etc.).

Examining cursors move around while the user is examining the page, not including time spent in 'reading' or 'action'.

Distinguishing Between Cursor Behaviors

We use a heuristic-based method to classify the different cursor behaviors. This method entails iteratively examining replays of the recorded interaction behaviors, deciding which behaviors belong to each category, classifying the behaviors using simple rules, and finally comparing the classified behaviors with the judged behaviors. The process is ad-hoc to develop a simple classification scheme that captures the essence of each behavior type.

We developed classification rules informed by watching replays of the interaction in a query session. The cursor is considered inactive if the user leaves it in one location (pausing it) while they examine the page. We define 'inactive' as the cursor staying still for at least one full second. The behaviors occurring when the cursor's position is active can be classified in three ways. As Claypool et al. noted, users may be using the cursor to help examine or read the page [10]. However, the cursor also serves the purpose of interacting with elements of the page or with the Web browser, typically by clicking on controls on the browser or Web page. We classify these 'action' behaviors as those occurring in the one second preceding a click. The remaining interactions were classified as either examining or reading. Reading was typically defined as users following the text horizontally (an observed pattern in Rodden et al. [28]), since the text in a snippet or advertisements goes from left to right. The movement to the right was not enough to classify as reading, since the cursor may be moved to the right for many reasons. We arrived at three rules for an interaction which if met, would label all

cursor positions within that timeframe as 'reading': the cursor could not have moved more than 50 px vertically (about 3 lines of text); the cursor must have moved at least 150 px to the right (the length of a handful of words); and the cursor must have moved back to the left at least 50 px. The remaining interactions, in which the cursor was neither inactive nor reading nor performing an action, were labeled as 'examining'. We therefore arrived at five cursor behavioral patterns (including clicking which was treated as an instantaneous action) as behaviors that warranted further analysis. We studied gaze-cursor alignment in these behaviors next.

Cursor Behavior's Effect on Alignment

For each of the five cursor behaviors, we computed the distance between gaze and cursor. Table 1 summarizes the proportion of time spent in each cursor behavior and the corresponding median distance between gaze and cursor. As expected, gaze and cursor are further apart when the cursor is inactive, since the eye is still roaming the SERP—233 px. Alignment is much closer when the cursor is being actively used to examine, read, or perform an action. The median distance when examining the page using the cursor is 167 px, while the alignment is closer when using the cursor to read—150 px. When the subject is moving the cursor to perform an action involving a click, alignment is extremely close—77 px, and even closer at the actual click—74 px. This agrees with findings from Hauger et al. that being in motion increases the alignment between cursor and gaze [20].

Behavior	Inactive	Examining	Reading	Action	Click
Total Time	58.8%	32.9%	2.5%	5.7%	-
Distance	233 px	167 px	150 px	77 px	74 px

Table 1. The median gaze-cursor distance for different cursor behaviors. The total time is summed across all subjects and search tasks.

During the study, the cursor was inactive a total of 18,554 seconds (58.8% of the time), representing time the subject may have been looking through the page without moving the mouse or just pausing for a few moments to read or think. This is more than the combined time of examining, reading, and performing an action on the page, meaning that a substantial period exists in which it is difficult to predict the gaze position. Anecdotally, some people believe they rarely use the cursor while examining Web pages. But in aggregate, a large portion of time is still spent actively moving the cursor, most of which does not produce an action.

Claypool et al. remarked, "Some users move the mouse while reading the window text or looking at interesting objects on the page, while others move the mouse only to click on interesting links" [10]. We quantified individual differences in gaze-cursor alignment and cursor behavior; Table 2 presents the duration of each cursor behavior and gaze-cursor alignment for each subject. The distances were macro-averaged over search tasks for each subject to negate the effect of a subject spending more time on certain search tasks.

Leaving the cursor idle from 50% to 79% of the time was common, like Subject 29 who left the cursor inactive during the majority of the time. Other subjects actively moved the

| Subject | Gender | Age | Inactive | | Examining | | Reading | | Action | | Click |
			Distance	Time	Distance	Time	Distance	Time	Distance	Time	Distance
1	F	46	188 px	68%	116 px	27%	45 px	0%	46 px	4.8%	49 px
2	M	56	201 px	65%	149 px	28%	133 px	0%	79 px	5.9%	73 px
3	F	35	165 px	47%	144 px	47%	167 px	2%	82 px	4.9%	75 px
4	M	47	162 px	61%	218 px	26%	214 px	6%	113 px	7.2%	106 px
5	F	31	246 px	45%	139 px	42%	134 px	6%	76 px	6.7%	89 px
6	M	44	175 px	59%	163 px	32%	211 px	1%	108 px	7.4%	101 px
7	M	57	144 px	60%	100 px	29%	93 px	6%	76 px	5.9%	81 px
8	F	47	309 px	63%	224 px	30%	143 px	3%	58 px	3.7%	43 px
9	F	50	262 px	56%	150 px	31%	146 px	8%	89 px	5.7%	100 px
10	M	59	180 px	65%	165 px	29%	209 px	1%	139 px	4.6%	126 px
11	M	50	247 px	50%	198 px	36%	256 px	4%	106 px	9.9%	81 px
12	F	39	236 px	29%	140 px	55%	125 px	5%	74 px	10.7%	77 px
13	F	45	177 px	55%	155 px	33%	204 px	4%	95 px	7.3%	71 px
14	M	50	206 px	49%	195 px	41%	177 px	2%	91 px	7.4%	93 px
15	F	38	294 px	68%	202 px	23%	168 px	2%	129 px	6.7%	88 px
16	F	26	152 px	61%	131 px	27%	118 px	2%	98 px	8.8%	84 px
17	F	46	218 px	55%	152 px	36%	145 px	4%	69 px	5.1%	65 px
18	M	50	245 px	58%	244 px	34%	251 px	2%	92 px	6.4%	116 px
19	F	50	185 px	55%	120 px	38%	142 px	2%	58 px	5.1%	46 px
20	M	39	370 px	44%	219 px	47%	140 px	3%	97 px	6.1%	70 px
21	M	38	146 px	39%	164 px	49%	139 px	1%	69 px	10.1%	59 px
22	M	47	216 px	69%	160 px	24%	173 px	3%	92 px	4.4%	71 px
23	F	46	218 px	50%	190 px	42%	196 px	1%	76 px	6.5%	64 px
24	F	48	224 px	61%	178 px	34%	89 px	0%	64 px	4.7%	47 px
25	F	54	228 px	56%	184 px	36%	173 px	1%	39 px	6.8%	26 px
26	F	44	166 px	32%	167 px	58%	116 px	1%	78 px	8.1%	109 px
27	F	28	161 px	58%	142 px	33%	112 px	0%	90 px	8.3%	52 px
28	F	51	259 px	56%	195 px	34%	242 px	2%	76 px	7.5%	72 px
29	F	41	218 px	79%	171 px	15%	148 px	1%	89 px	5.7%	119 px
30	M	48	279 px	50%	138 px	37%	87 px	5%	42 px	7.7%	38 px
31	M	59	192 px	58%	175 px	34%	162 px	3%	152 px	4.5%	133 px
32	F	53	236 px	61%	199 px	35%	77 px	0%	52 px	3.3%	77 px
33	M	60	418 px	62%	305 px	27%	319 px	3%	69 px	8.0%	105 px
34	F	47	245 px	59%	214 px	36%	160 px	1%	44 px	4.2%	34 px
35	M	44	187 px	73%	133 px	22%	167 px	1%	68 px	4.7%	85 px
36	F	37	290 px	76%	231 px	20%	147 px	1%	58 px	2.9%	56 px

Table 2. Basic demographic information for each subject, the proportion of time spent performing each cursor behavior, and the average distance between gaze and cursor while performing that cursor behavior. Clicks are considered instantaneous events and therefore have no duration.

cursor while examining the page; Subject 12 spent 55% of their time examining the page with the cursor, and a mere 29% of their time idling the cursor. Reading behavior comprised 2% or less of total search time for more than half the subjects (22 of 36), who exhibited nearly no reading behavior (per the definition of this behavior that we used). At the other end of the spectrum, Subject 9 spent 8% of their time reading with the cursor. Turning to gaze-cursor distance, some subjects had poor alignment (Subject 33) when not performing an action. While in the aggregate, subjects had stronger alignment when the cursor was active, Subject 18 had essentially unchanged alignment between inactive, examining, and reading behaviors. These differences show that individuals vary substantially in their cursor behavior usage and gaze-cursor alignment. We initially thought that perhaps the variation in users' gaze-cursor alignment could be explained by choice of cursor behavior, but for each cursor behavior, gaze-cursor alignment still varied substantially among our subjects.

PREDICTING GAZE

So far, we have seen the effects of user, query, dwell time, cursor behavior, and future cursor positions on gaze-cursor alignment. These features can provide guidance about whether the alignment is stronger or weaker. Predicting the strength of alignment is useful in determining the confidence of the gaze position from only interaction features—without using an eye-tracker. We now focus on a more aggressive predic-

tion: to find the position of the gaze using only interaction features, rather than how confident we are that the cursor position approximates the gaze position. To do this, we extract four types of features from the interaction data, aiming to select features that seemed to influence the gaze-cursor alignment, as informed by our earlier analysis.

Features

At each time t that we want to predict the gaze position, we have the **cursor position** at that time, represented by a tuple (x, y). The cursor position alone is suggested for approximating gaze position in some prior literature and current Web analytics services. This approach is the baseline against which we will compare our performance, but we also use cursor position as a feature in our gaze prediction model.

Cursor behavior has a strong effect on gaze-cursor alignment. Both our earlier analysis and a study by Hauger et al. [20] showed that active cursors are better aligned with gaze than inactive cursors. To generalize this, we use the idle time following the last movement before t as the **behavior** feature, representing activity level.

Each recorded interaction on the Web page has a corresponding timestamp, allowing us to deduce the length of time since the SERP has loaded. Our analysis has shown that the time since the SERP loaded influences the gaze-cursor alignment. This time ($\text{current}_t - \text{pageload}_t$) is the **dwell** feature, which

we incorporate into our model. Guo and Agichtein also use this feature to predict gaze-cursor alignment [19].

Our analysis showed that for every subject in our study, the cursor position lagged behind the gaze position, since there is a stronger correlation between a future cursor position and the current gaze position. We refer to the most likely later cursor position for the current gaze as the **future** feature. The future cursor positions were only used if the last movement was within 10 seconds of the target future time.

We also used the interaction effect of cursor position with dwell and cursor position with behavior, since we know that behavior and dwell affect alignment. The user or query were not treated as features for two reasons. First, in practical situations, a user's gaze data is not available to train the model. Second, there is unlikely to be enough gaze data for most queries to train the model, and in our analysis described earlier in the paper, we found that query only has a modest effect on gaze-cursor alignment. Our current features are all global and so require fewer training examples in the form of eye-tracking data (i.e., actual gaze positions).

Experiment

We predict the subject's gaze position using a linear model of interaction features. The ground truth is the gaze position measured by the eye-tracking system. The x- and y-coordinates of the gaze position were predicted separately. To compute the weights (coefficients) for each feature, we performed a multiple linear regression. Figure 6 illustrates the value of gaze prediction in an example query session with cursor positions, gaze positions, and predicted gaze positions overlaid on the SERP.

The model for the regression for the x-coordinate is:

$$g_x \sim c_x + \log(t_d) + \log(t_m) + c_x \times \log(t_d) + c_x \times \log(t_m) + f_x \quad (2)$$

where t_d is the dwell time, t_m is the time since a movement, c_x is the x-coordinate of the cursor position, g_x is the x-coordinate of the gaze position, and f_x is the most likely x-coordinate of the gaze based on future cursor positions. The regression for the y-coordinate was similar, but substituting x for y in Equation 2.

The evaluation was a 36-fold cross-validation, where each fold was an individual subject; this is essentially a leave-one-out evaluation for each subject. By testing each subject separately with the training data of all 35 other subjects, we achieve a practical method of predicting the gaze position for users that we have not seen before (which would be the typical case in a real deployment). Essentially, we ran a multiple linear regression on the gaze and interaction data of 35 subjects to compute weights (coefficients), then used those coefficients to predict gaze data from just the interaction data for the test subject. This process was repeated for every subject being the test subject. For each subject we tested, we computed the RMSE for their predicted gaze with and without the future feature, as well as the RMSE for just the cursor positions. Without using the future feature, there is a potential application for real-time gaze prediction, while with future cursor positions, the gaze prediction must be performed off-line (after the query session, e.g., for analytics).
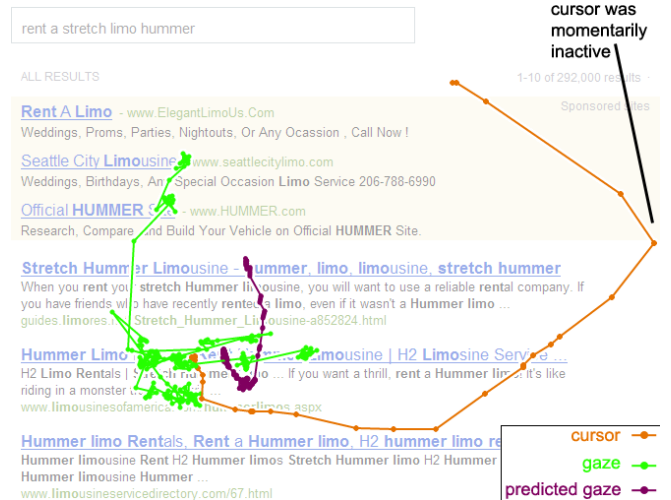

Figure 6. The cursor position (orange), the gaze position predicted by the linear regression model (purple), and the gaze position as determined by the eye-tracker (green) are drawn over the SERP presented to the subject following the query "rent a stretch limo hummer" from our study. The figure omits the right and left columns of the SERP.

Table 3 shows the results of the prediction experiment. When using only the cursor position for prediction, the distance RMSE is 236.6 px. But using a multiple linear regression with cursor position, behavior, and dwell time, the predicted gaze position is significantly better—186.3 px, a 21.3% decrease in RMSE. Adding future cursor data from that query session to the model reduces RMSE further to an overall 23.5% decrease in RMSE compared to just using the cursor position. The RMSEs in the x- and y-coordinates alone were similarly improved by the linear model. An ANOVA shows that the RMSE differed significantly between gaze-cursor alignment and alignment between gaze and predicted gaze along the x-axis ($F(2, 105) = 59.72, p < 0.001$) and Euclidean distance ($F(2, 105) = 41.31, p < 0.001$).

The Lindeman, Merenda and Gold (LMG) metric [24] (R^2 partitioned by averaging over orders) determined the relative importance of the features for predicting the x-coordinate. They were in descending order of importance: $\log(t_d)$, c_x, f_x, $c_x \times \log(t_m)$, $\log(t_m)$, and $c_x \times \log(t_d)$; the relative importance of the features for predicting the y-coordinate in descending order: c_y, f_y, $\log(t_d)$, $\log(t_m)$, $c_y \times \log(t_m)$, and $c_y \times \log(t_d)$.

	RMSE_x	RMSE_y	RMSE_d
Cursor (baseline)	185.0 px	145.0 px	236.6 px
Predicted Gaze with Cursor+Behavior+Dwell	125.2 px	137.1 px	186.3 px
Predicted Gaze with Cursor+Behavior+Dwell+Future	125.1 px	129.9 px	181.1 px

Table 3. The computed accuracies in cross-validation evaluations of estimating the gaze position using the cursor position, the cursor position along with behavior and duration using multiple linear regression, and the cursor position along with behavior, duration, and future cursor positions using multiple linear regression. The accuracy is measured by root-mean-square error for the x-axis, y-axis, and Euclidean distance.

DISCUSSION

Our study has shown that gaze-cursor alignment is situational, as it depends on the time spent on the page, personal browsing habits, and a user's current cursor behavior (inactive, examining, reading, action). An experiment showed that a model using these features could predict the subject's gaze significantly better than using the cursor position alone. These findings have implications for using large-scale cursor data more effectively, which has already been demonstrated to be efficiently obtainable at scale [2, 22].

Our findings suggest that for certain circumstances, it may be possible to predict the actual gaze position on SERPs using only cursor features. This extends previous work [19] which attempted the binary prediction task of whether we can be confident of gaze-cursor alignment within a threshold. Using a linear model, we show that there is room for improving gaze prediction by using other factors on top of solely the cursor position. Our predictive model reduces the RMSE by 60 px in the x-direction and is 15 px more accurate in the y-direction. 15 px is around two lines of text on a SERP, and 60 px is around 10 characters of that text. These gains could be significant for differentiating between engagement with regions on the SERP, especially when the cursor is near the boundary of two or more regions. However, since the linear regression fits using least squares, the predicted gaze tends to be conservative and often stays around the center of the screen. This leads us to believe that gaze prediction may be improved further by a more complex combination of the features, perhaps in non-linear models, since we intuit that interaction effects of time, behavior, and cursor position exist.

The gaze prediction results in Table 3 suggest a counterintuitive finding. While several past studies [19, 22, 27] agree that the cursor is better correlated with the gaze in the y-direction than the x-direction, the x-coordinate of the gaze is easier to predict than the y-coordinate. This subverts the expectation that better correlation leads to better prediction for gaze. This may be an artifact of our model, but perhaps left-right eye movement is less surprising than up-down movement. Scrolling may cause the cursor to move down relative to the page, making it a good estimate of vertical attention, whereas most SERPs require little horizontal scrolling.

Cursor tracking, deemed the "poor man's eye tracker" [12], may approximate gaze tracking without the eye tracker depending on the accuracy required. Although cursor features would allow us to model many aspects of user attention *in situ* as they browse the Web from home, they cannot completely replace gaze. For example, eye-gaze fixation is a positive signal of interest because the user pays more attention to that position, but prolonged cursor fixation may not be since given our findings in this study, the user's attention is probably elsewhere. Here we elected to focus on understanding different cursor behaviors rather than gaze fixations to support cursor-tracking applications that can be remotely deployed on a Web site. More work is needed to study in more detail the relationship between cursor and gaze fixations, especially to determine if and when there are cases in which cursor fixations can be reliably interpreted as attention.

Bringing subjects into an eye-tracking lab creates inherent limitations to our study. Subjects may behave differently in the lab with a camera monitoring their gaze than in a natural setting. They are unlikely to take breaks in the lab or multitask by doing other things at the same time, factors which may affect cursor behavior outside of the lab. We also gave them artificial search tasks; while we tried to ground the search tasks in realistic information needs, the tasks might not be a representative sample. Additionally, though SERPs provide a controlled environment for our studies (and there are already several applications if we focus on that domain alone), more work is also needed to generalize this research beyond SERPs to any Web page. Previous work has already shown that cursor is less aligned with gaze on non-SERP pages [22], so the prediction task on such pages may be more challenging and is reserved for future work.

The results show that we predicted gaze more accurately when we used past and future cursor movement data, rather than only past data. Using past data alone would allow us to build applications that could respond to user attention in real-time, such as a focus-plus-context view of the SERP, in which the context could update dynamically with new content based on where the user had already attended during their engagement with the SERP. More valuable however, is inferring gaze positions after the user has left the SERP. This would allow us to accurately model where on the SERP the user examined and use this data for applications such as building richer searcher models (improving on existing models, e.g., [8, 16]), usability assessments [1, 2], or profiling users [2].

CONCLUSION

We have found that user, time, and search task (to a lesser extent) each contribute to the variation in gaze-cursor alignment. The gaze and cursor positions are also better aligned when the gaze position is compared to a future cursor position. Furthermore, by distinguishing between five different cursor behaviors—inactive, examining, reading, action, and click—we get a better idea of the strength of alignment. In fact, we have been able to improve upon using cursor position alone to predict the gaze position by using several cursor features. Cursor movements, scrolling, and other client-side interactions are easy to collect at scale, which many Web analytics services offer to do. But claiming that the cursor approximates the gaze is misguided—as we have shown, this is often not the case depending on time and behavior. Instead, it is important to predict the real location of the attention when an eye-tracker is unavailable.

REFERENCES

1. Arroyo, E., Selker, T., and Wei, W. Usability tool for analysis of web designs using mouse tracks. In *CHI Extended Abstracts* (2006), 484–489.

2. Atterer, R., Wnuk, M., and Schmidt, A. Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction. In *Proceedings of WWW* (2006), 203–212.

3. Aula, A., Majaranta, P., and Räihä, K.-J. Eye-tracking reveals the personal styles for search result evaluation. In *Proceedings of INTERACT* (2005), 1058–1061.

4. Buscher, G., Cutrell, E., and Morris, M. R. What do you see when you're surfing?: using eye tracking to predict salient regions of web pages. In *Proceedings of CHI* (2009), 21–30.

5. Buscher, G., Dengel, A., and van Elst, L. Eye movements as implicit relevance feedback. In *CHI Extended Abstracts* (2008), 2991–2996.

6. Buscher, G., Dumais, S. T., and Cutrell, E. The good, the bad, and the random: an eye-tracking study of ad quality in web search. In *Proceedings of SIGIR* (2010), 42–49.

7. Buscher, G., van Elst, L., and Dengel, A. Segment-level display time as implicit feedback: a comparison to eye tracking. In *Proceedings of SIGIR* (2009), 67–74.

8. Chapelle, O., and Zhang, Y. A dynamic bayesian network click model for web search ranking. In *Proceedings of WWW* (2009), 1–10.

9. Chen, M. C., Anderson, J. R., and Sohn, M. H. What can a mouse cursor tell us more?: correlation of eye/mouse movements on web browsing. In *CHI Extended Abstracts* (2001), 281–282.

10. Claypool, M., Le, P., Wased, M., and Brown, D. Implicit interest indicators. In *Proceedings of IUI* (2001), 33–40.

11. Cole, M. J., Gwizdka, J., Liu, C., Bierig, R., Belkin, N. J., and Zhang, X. Task and user effects on reading patterns in information search. *Interacting with Computers 23*, 4 (2011), 346 – 362.

12. Cooke, L. Is the mouse a poor man's eye tracker? In *Proceedings of STC* (2006).

13. Cutrell, E., and Guan, Z. What are you looking for?: an eye-tracking study of information usage in web search. In *Proceedings of CHI* (2007), 407–416.

14. Dumais, S. T., Buscher, G., and Cutrell, E. Individual differences in gaze patterns for web search. In *Proceedings of IIiX* (2010), 185–194.

15. Goecks, J., and Shavlik, J. Learning users' interests by unobtrusively observing their normal behavior. In *Proceedings of IUI* (2000), 129–132.

16. Guo, F., Liu, C., Kannan, A., Minka, T., Taylor, M., Wang, Y.-M., and Faloutsos, C. Click chain model in web search. In *Proceedings of WWW* (2009), 11–20.

17. Guo, Q., and Agichtein, E. Exploring mouse movements for inferring query intent. In *Proceedings of SIGIR* (2008), 707–708.

18. Guo, Q., and Agichtein, E. Ready to buy or just browsing?: detecting web searcher goals from interaction data. In *Proceedings of SIGIR* (2010), 130–137.

19. Guo, Q., and Agichtein, E. Towards predicting web searcher gaze position from mouse movements. In *Proceedings of CHI* (2010), 3601–3606.

20. Hauger, D., Paramythis, A., and Weibelzahl, S. Using browser interaction data to determine page reading behavior. In *Proceedings of UMAP* (2011), 147–158.

21. Hijikata, Y. Implicit user profiling for on demand relevance feedback. In *Proceedings of IUI* (2004), 198–205.

22. Huang, J., White, R. W., and Dumais, S. No clicks, no problem: using cursor movements to understand and improve search. In *Proceedings of CHI* (2011), 1225–1234.

23. Ianniello, J. Time delay estimation via cross-correlation in the presence of large estimation errors. *IEEE Trans. Acoust., Speech, Signal Process. 30*, 6 (1982), 998–1003.

24. Lindeman, R., Merenda, P., and Gold, R. *Introduction to Bivariate and Multivariate Analysis*. Scott Foresman, Glenview IL, 1980, 119.

25. Liu, C.-C., and Chung, C.-W. Detecting mouse movement with repeated visit patterns for retrieving noticed knowledge components on web pages. *IEICE - Trans. Inf. Syst. E90-D*, 10 (October 2007), 1687–1696.

26. Reichle, E. D., Pollatsek, A., and Rayner, K. E-z reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research 7*, 1 (2006), 4 – 22.

27. Rodden, K., and Fu, X. Exploring how mouse movements relate to eye movements on web search results pages. In *SIGIR Workshop on Web Information Seeking and Interaction* (2007), 29–32.

28. Rodden, K., Fu, X., Aula, A., and Spiro, I. Eye-mouse coordination patterns on web search results pages. In *CHI Extended Abstracts* (2008), 2997–3002.

29. Shapira, B., Taieb-Maimon, M., and Moskowitz, A. Study of the usefulness of known and new implicit indicators and their optimal combination for accurate inference of users interests. In *Proceedings of SAC* (2006), 1118–1119.

30. Teevan, J., Dumais, S. T., and Liebling, D. J. To personalize or not to personalize: modeling queries with variation in user intent. In *Proceedings of SIGIR* (2008), 163–170.

31. Torres, L. A. L., and Hernando, R. V. (smt) real time mouse tracking registration and visualization tool for usability evaluation on websites. In *Proceedings of IADIS WWW/Internet* (2007), 187–192.

32. Ware, C., and Mikaelian, H. H. An evaluation of an eye tracker as a device for computer input. In *Proceedings of CHI* (1987), 183–188.

33. Weber, I., and Castillo, C. The demographics of web search. In *Proceedings of SIGIR* (2010), 523–530.