Large-Scale Analysis of Individual and Task Differences in Search Result Page Examination Strategies

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

Ryen W. White and Susan T. Dumais
Microsoft Research
Redmond, WA 98052
{ryenw, sdumais}@microsoft.com

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

ABSTRACT
Understanding the impact of individual and task differences on search result page examination strategies is important in developing improved search engines. Characterizing these effects using query and click data alone is common but insufficient since they provide an incomplete picture of result examination behavior. Cursor- or gaze-tracking studies reveal richer interaction patterns but are often done in small-scale laboratory settings. In this paper we leverage large-scale rich behavioral log data in a naturalistic setting. We examine queries, clicks, cursor movements, scrolling, and text highlighting for millions of queries on the Bing commercial search engine to better understand the impact of user, task, and user-task interactions on user behavior on search result pages (SERPs). By clustering users based on cursor features, we identify individual, task, and user-task differences in how users examine results which are similar to those observed in small-scale studies. Our findings have implications for developing search support for behaviorally-similar searcher cohorts, modeling search behavior, and designing search systems that leverage implicit feedback.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval–*search process, selection process*.

General Terms
Experimentation, Human Factors.

Keywords
Rich interaction logging, individual differences, task differences

1. INTRODUCTION
Better understanding how users interact with Web search engines is important for improving the search experience. The information science community has studied individual differences in search strategies, tactics, and performance, and identified important factors such as prior experience, gender, age, cognitive styles, interface design, and domain expertise [1][4][30] that influence search strategies and task performance. Most of these studies are conducted in a controlled laboratory setting which limits the number of participants and the naturalness of the tasks selected for study. Thus it is unclear the extent to which these findings generalize to the wide diversity of searchers' tasks seen in Web search settings.

WSDM'12, February 8–12, 2012, Seattle, Washington, USA.

The Web provides unprecedented opportunities to evaluate alternative design, interaction, and algorithmic methods at scale and *in situ* with actual people doing their own tasks in their own environments [22]. Studies of searcher engagement with search engine result pages (SERPs) focus primarily on search result clickthrough behavior. These studies provide insights regarding the order in which results are clicked. However, they fail to capture behaviors that do not lead to clicks (e.g., which items are attended to, in what order, etc.) or subjective impressions.

There are two main ways to capture detailed behavioral patterns in search: gaze tracking and mouse cursor logging. Gaze tracking studies can provide more detailed insights about how visual attention is distributed on the SERP (and subsequent pages). However these studies are typically conducted in laboratories using a small number of participants with assigned tasks (e.g., [9][15]), with summaries of gaze behavior aggregated across participants and tasks. Some studies examined individual and task differences in gaze patterns, and found individual differences in the strategies with which users inspect results [12], and different clusters of users who exhibit similar result examination behaviors [3][12]. Others found that the type of search task (informational vs. navigational) influenced task completion time and time spent reviewing documents [9][27]. Gaze tracking can provide valuable insights but the technology is expensive and needs calibration, meaning that it does not scale well to non-laboratory settings. A new technique, *ViewSer,* allows for approximate tracking of gaze at much larger scale. It does so by blurring the SERP and only revealing the region proximal to the mouse pointer in more detail [23]. However, this method influences the SERP's visual presentation which likely also affects users' SERP examination strategies.

An alternative to gaze tracking is mouse cursor tracking. Recent research has shown that cursor movements correlate with eye gaze [5][17][28][29], and may therefore be an effective indicator of user attention. Small-scale laboratory studies have observed participants making many uses of the cursor on SERPs beyond hyperlink clicking [2][25]. These uses include moving the cursor as a reading aid, using it to mark interesting results, using it to interact with controls on the screen (e.g., buttons, scroll bars), or simply positioning the cursor so that it does not occlude Web page content. However, these studies were in small-scale laboratory settings which limit what inferences can be made about more naturalistic search behavior. Cursor tracking provides an efficient and unobtrusive way to track mouse movement behavior and can be deployed at scale [18]. We believe that rich cursor tracking data affords a detailed analysis of user, task and user-task differences that is not possible with current evaluation methodologies.

In the research reported in this paper, we use rich cursor logs gathered from a deployment on the Bing commercial Web search engine to better understand individual and task effects on SERP interaction. In addition to tracking mouse behaviors such as cursor movements and clicks, we also logged the location of all areas of interest (to enable accurate assignment of movements to SERP elements), viewport size, scrolling activity, and text selections. Rich data of this type were captured for over 1.8 million queries using a methodology similar to that proposed by Huang *et al.* [18]. This provides sufficient data to allow us to investigate the effects of user and task differences, and interactions between them, and reach conclusions which are potentially generalizable. As we show in our analysis there are distinct user clusters exhibiting specific SERP interaction strategies that can be observed from these data, and are particularly apparent when we also consider the effect of search task on users' SERP examination behaviors.

We make the following contributions with this research:

- Gather and use rich cursor interaction log data on a Web scale;
- Automatically identify distinct users clusters at scale based on SERP examination behaviors, and relate these clusters to findings from previous smaller-scale user studies;
- Study the effect of task type (navigational vs. non-navigational) and consider their impact on the user clustering, and;
- Propose design implications based on our behavioral clustering.

The remainder of this paper is structured as follows. Section 2 describes related work on individual differences in search behaviors, past work on the effect of search task on search behavior, and previous work on gaze- and cursor-tracking. Section 3 describes the large scale cursor tracking data, including the methodology used to gather the data and summary statistics on SERP interaction. Section 4 describes the features that we extracted from the data, the findings of our analysis of user differences, task differences, and any interactions between them. We discuss findings and their implications in Section 5, and conclude in Section 6.

2. RELATED WORK

Three lines of prior research are related to the work described in this paper: (i) examining individual differences in search behaviors and strategies, (ii) studying the relationship between search task and search behavior, and (iii) characterizing how people interact with SERPs using gaze- and cursor-tracking studies.

Saracevic summarized the long history in information science of understanding how individual differences influence search strategies and task performance [30]. Allen [1] showed that cognitive styles, specifically field dependence, influences search task performance. Ford *et al.* [13] showed that a variety of individual differences including cognitive style, search experience and age influence both search strategies and task outcome. Bhavnani [4] and Thatcher [33] examined how search behavior varies with domain and search expertise. Bhavnani showed that domain knowledge influences the choice of search strategies and search success. Thatcher observed search differences related to Web experience with experts using more known URL addresses and parallel strategies. These studies provided very detailed modeling of searcher behaviors, often coupled with survey data to better understand motivations, but are laboratory studies involving small numbers of searchers and tasks. At the other end of the spectrum, large-scale log analyses examined the relationship between search expertise (White and Morris [34]) and domain expertise (White *et al.* [36]) on Web search behaviors. White *et al.* found that domain experts are more successful than novices (in the domain of their expertise) and achieve this success using different vocabulary, sites and broader search strategies. Using a different strategy of clustering users with similar behavioral patterns (rather than using known cognitive, skill or demographic differences), White and Drucker [35] identified two general types of Web searchers: navigators (with very consistent search and browsing patterns) and explorers (with much more varied search and browsing patterns).

There has also been research on the relationship between search tasks and search behavior. Using a diary study, Byström and Järvelin [6] looked directly at the impact of task complexity on user search behavior, examining the relationships between task complexity, information types, and information sources. They showed that as task complexity increased, users needed more sources of information, more domain information and more problem solving information, were less likely to predict the types of information they needed, and were more dependent upon experts to provide useful information. Kellar *et al*. [21] used a field study to examine four task types: fact-finding, information gathering, browsing, and transactions, and examined how users interacted across them as they navigated the Web. They showed that the information gathering task was most complex: participants spent the most time completing it, viewed more pages, and used browser functionality most heavily. Liu *et al*. [26] investigated user behaviors associated with different task types in a controlled laboratory experiment. They varied tasks on different dimensions: complexity, product, goal, and level. Their results indicate differences in search behaviors associated with different task characteristics, including task completion time, the time to assess document utility, and eye fixations. They further suggest that these implicit behaviors could be indicative of task type.

In addition to Liu *et al.*, others have used eye-tracking to provide detailed quantitative analyses of the distribution of gaze as people perform search tasks (e.g., [5][9][15][23][27][28][29]). Since eye gaze position is highly correlated with visual attention, these studies provide rich insight into what people are attending to as they interact with SERPs. Several studies characterized how visual attention is distributed over the search results [15][23][27], or between search results and advertisements [5]. Guan and Cutrell [15] and Lorigo *et al.* [27] found differences in search time and examination patterns for informational vs. navigational tasks. Cole *et al.* [8] identified differences in reading patterns associated with different task characteristics and page types. Aula *et al.* [3] identified two general patterns that people used in examining search results: exhaustive evaluators (54% of the participants who looked at more than half of the visible results for more than half the tasks) and economic evaluators (46% of the participants). Dumais *et al.* [12] performed a similar analysis of search behavior using more complex result pages that included both organic results and advertisements. They found three general groups of searchers – exhaustive (32%), economic with a focus on results (39%), and economic with a focus on advertisements (29%). Although gaze-tracking provides detailed insight into search behavior, it requires calibration, is laboratory based, and does not scale well to the wide range of tasks and users observed in Web search.

To address these issues, cursor-tracking has recently been used to examine search behavior. Initial studies established a close correspondence between eye gaze and cursor position [5][18][28][29]. More recent studies have looked at ways in which cursor movements can be used to understand search behavior. In small-scale studies, Guo and Agichtein used cursor movement to predict query intent [16], and to predict gaze position [17]. In another small-scale study, Rodden *et al.* [29] identified four general uses of the cursor in Web search – neglecting the cursor while reading, using the cursor as a reading aid (either horizontally or vertically), and using the cursor to mark interesting results. In a larger-scale study, Huang *et al.* [18] summarized how cursor activity (including clicks on hyperlinks, clicks on non-hyperlinks, and search result snippet hover behavior) related to Web search behavior. They also showed how cursor activity could be used to estimate the relevance of search results and to differentiate between good and bad SERP abandonment. Rather than tracking the mouse cursor at scale, Lagun and Agichtein [23] presented a scalable method to estimate gaze position by blurring the SERP and only revealing a region proximal to the mouse cursor. They found that result viewing and clickthrough patterns agree closely with unrestricted viewing of results, as measured by eye-tracking.

The research presented in this paper extends the previous work presented in this section in several ways. First, we describe a large-scale analysis of cursor behavior (including clicks, hovers, text selection and cursor trails) on different regions of a search results page. Second, we use these implicit signals of user engagement with search result pages to cluster individuals with similar patterns of behavior. Finally, we examine how user, task and user × task interactions influence search behavior.

We begin by describing the interaction log data used in this study.

3. INTERACTION LOG DATA

We recorded interaction data directly on the SERP of the Bing commercial Web search engine. Log data were gathered over a period of 13 days between May 26, 2011 and June 7, 2011 during an external experiment on a small fraction of user traffic.[1] In the following, we describe our logging methods and provide an initial overview of the data gathered.

3.1 Methodology

To record user interactions with the SERP at scale without the need to install any browser plugins, we used an efficient and scalable approach similar to that developed by Huang *et al.* [18]. As such, JavaScript-based logging functions were embedded into the HTML source code of the SERP. To obtain a detailed understanding of user interactions with the SERP, we recorded information on mouse cursor movements, clicks, scrolling, text selection events, focus gain and loss events of the browser window, as well as bounding boxes of several areas of interest (AOIs) on the SERP and the browser's viewport size. Combining these data sources enabled us to develop a rich picture of how searchers engaged with the SERP, something not previously possible at scale.

When logging any additional type of user interaction data beyond clickthrough, a tradeoff has to be made between: (i) level of detail (e.g., temporal and spatial resolution), (ii) the impact of any additional JavaScript code on page load time, and therefore the user experience, which can be sensitive to even small increases in load time, and (iii) the amount of data transferred (and hence bandwidth consumed) between the client and the remote server as well as log volume created on the backend server.

We now describe in more detail the fields that are recorded in our log data and the methods used to record them.

3.1.1 Mouse Cursor Position

The JavaScript function for logging mouse cursor positions checked the cursor's x- and y-coordinates relative to the top-left corner of the SERP every 250 milliseconds. Whenever the cursor had been moved more than eight pixels away from its previously logged position, its new coordinates were sent to the remote Web server. Eight pixels correspond to approximately the height of half a line of text on the SERP. We used this approach rather than recording every cursor movement since we wanted to minimize the data gathered and transmitted so as to not adversely affect the user experience with delays associated with log data capture and data uploads to the remote server. Since cursor tracking was relative to the document, we captured cursor alignment to SERP content regardless of how the user reached that position (e.g., by scrolling or keyboard). Therefore this approach did not constrain other behaviors such as scrolling or keyboard input.

3.1.2 Mouse Clicks

Mouse clicks were recorded using the JavaScript *onMouseDown* event handling method. Thus, the backend server received log entries with location coordinates for *every* mouse click, including clicks that occurred on a hyperlink as well as those that occurred elsewhere on the page (even on white space containing no content). To identify clicks on hyperlinks and differentiate them from clicks on inactive page elements, we also logged unique hyperlink identifiers embedded in the SERP.

3.1.3 Scrolling

We also recorded the current scroll position, i.e., the y-coordinate of the uppermost visible pixel of the SERP in the browser viewport. This coordinate was checked three times per second and was recorded whenever it had changed by more than 40 pixels compared to the last logged scrolling position. Forty pixels correspond to the height of about two lines of text. From this coordinate we gain a number of insights into scrolling behavior, including whether the user scrolled up or down, and the maximum scroll depth, to understand how far down the SERP the user scrolled.

3.1.4 Text Selections

Searchers may select text for a number of reasons, including to copy-and-paste to another application or to issue a new query to a search engine. Using browser-specific JavaScript functionality, we could identify when text selections occurred and could also determine the bounding box of the immediately surrounding HTML element inside which the selection occurred. For every text selection we recorded the coordinates of the upper left corner of the determined element's bounding box. The actual contents or the exact position of the selected text were not recorded.

3.1.5 Viewport Size

The width and height of the browser viewport in pixels at SERP load time were also logged. Cases where the browser window was resized during interaction with the SERP were not accounted for.

3.1.6 AOI Positions

Simply logging the text of what was displayed on the SERP is insufficient for reconstructing its layout since SERPs vary per query (depending on whether answers are shown, etc.), font sizes, and other browser preferences. To reconstruct the exact SERP layout as it was rendered in the user's browser, we recorded the positions and sizes of AOIs. The specific AOIs that we were interested in were: (i) top and bottom search boxes, (ii) left rail and its contained related searches, search history, and query refinement areas, (iii) mainline results area and its contained result entries, including advertisements and answers, and (iv) right rail. Some of these AOIs are visualized overleaf in Figure 1.

For each AOI bounding box, we determined and logged the coordinates of its upper left corner as well as its width and height in pixels. Using this information, we could later map cursor positions, clicks, and text selections to specific AOIs.

Before describing our analysis of user and task differences, we first provide some summary statistics on the data set gathered.

[1] Note that these data differ from those used by Huang *et al.* [18], in that they are from external users and not Microsoft employees.

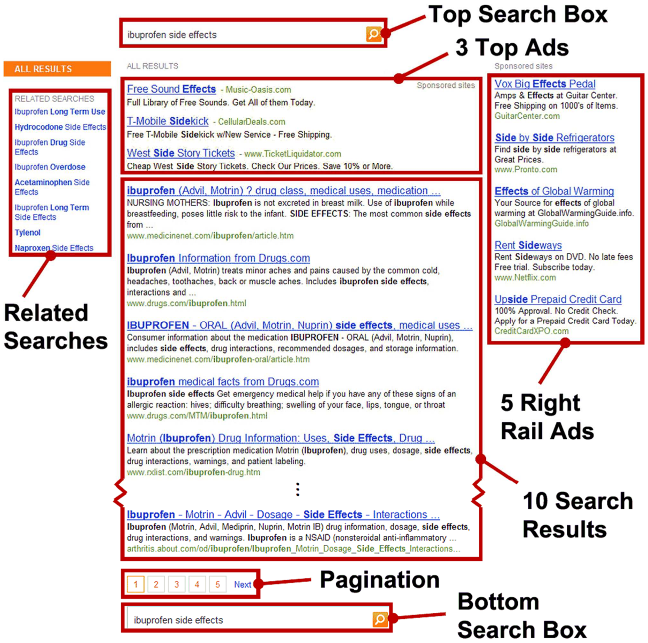


Figure 1. Recorded AOIs on an example SERP.

3.2 Summary Statistics

As described in the previous subsection, the raw interaction events comprised cursor positions, clicks, window scrolls, and text selection events. In total, our data set contained over 1.8 million queries. The average amount of time on the SERP was 47.2 seconds (median=5 seconds). Other summary measures specific to the different types of data logged include:

Cursor: The average speed of the cursor was 172.3 pixels (px) per second (median=123.9 px/sec). The total length of the cursor trail depicting the path that the trail follows on the SERP is 1467.3 px (median=630.7 px), and the trail changes direction (defined in terms of the compass directions North, South, East, and West) almost four times in the course of the trail (mean=3.8, median=1). Finally, users hover over multiple result captions (mean=2.6, median=2), even for navigational queries when a single search result will suffice. This behavior pattern has been observed in studies of eye tracking [9], as well as previous work on cursor tracking [18].

Clicks and text selections: Mouse clicks were collected regardless of whether they were on a hyperlink or on other regions of the page. 64.7% of all clicks were hyperlink clicks and 35.2% were non-hyperlink clicks, which are missed by traditional click-based instrumentation methods. Non-hyperlink clicks were on empty regions of the page (58.1% of all SERPs), on controls (38.8% of all SERPs), and on the upper or lower search boxes (11.4% of all SERPs). Text selections occurred on 1.9% of all SERPs.

Scrolling: Window scrolling is a client-side interaction that is rarely captured in the context of Web search. Of the queries in our set, 29.7% contained at least one scroll event. 61.8% of logged interaction sequences for a query ended on a downwards scroll. As expected, there were more downward scrolls than upward scrolls, and the majority of scrolled queries (54.8%) comprised only downward scrolls. This suggests that most queries do not result in the user returning to the top of the SERP to examine search results that may be hidden following scrolling.

We now extract features from the events and cluster them to analyze the relationships between search behavior, users, and tasks.

4. USER AND TASK DIFFERENCES

The data that we gathered allows us to study how user and task differences impact search result page examination behavior. This is important for better understanding how users engage with search engines and informing the design of new kinds of search support tailored to observed strategies. Using the cursor data described in Section 3.1, we developed summary and composite features to characterize search behavior.

4.1 Feature Extraction

There were four main classes of SERP interaction features: (i) cursor, describing features related to movement of the mouse cursor; (ii) click, related to clicks (both hyperlink and otherwise); (iii) scrolling, describing scrolling behavior (using scrollbar, scroll wheel on mouse, or keyboard commands); and (iv) other features such text selections and interactions with specific SERP features particular to the search engine on which this study was performed. These features are aggregated at the level of a single SERP.

4.1.1 Cursor

We computed a number of cursor-related features based on cursor movements, positions, and dwells, which fall into four groups:

Trails: These features are derived from the recorded cursor movement trails on the SERP and include trail length, trail speed, trail time, total number of cursor movements, and summary values (average, median, standard deviation) for single cursor movement distances and cursor dwell times in the same position, etc. We also created features for the total number of mouse movements and the total number of times that the cursor changed direction in the trail.

Hovers: We recorded total hover time on the SERP. Since we recorded the coordinates of the AOIs we were also able to associate cursor movements with particular SERP elements (see Figure 1). This allowed us to represent the total hover time on inline answers (e.g., stock quotes, news headlines, etc.), in the lower and upper search box on the result page, in the left rail (where search support such as query suggestions and search history would usually be shown), in the right rail (where advertisements would usually be shown), and in the algorithmic results. We also computed the total amount of time that the mouse cursor was idle on the SERP.

Result Inspection Patterns: We computed features summarizing how users inspected the search results, including the total number of search results that users hovered over, the average result hover position, and the fraction of the top ten results that were hovered. We also created features of the linearity of scanning over search results using the same basic definitions introduced by Lorigo *et al.* [27]. We obtain a numbered *scan sequence* by assigning numbers to all top ads (i.e., -2, -1, 0) and organic results (i.e., 1, 2, …, 10) and using these numbers to describe the scan path. The *minimal scan sequence* is obtained by removing repeat visits to a result entry. We had binary features describing whether the scan sequence and the minimal scan sequence were strictly linearly increasing (e.g., 1, 2, 3, …) meaning that users were traversing the search results linearly, as has been suggested previously [19].

Reading Patterns: We used sequences of cursor actions to identify reading-with-mouse behavior (frequent left-right-left movements) which is one of the behaviors identified by Rodden *et al.* [29]. We encoded mouse moves (specified by two adjacent cursor movement events) with the symbols “N”, “E”, “S”, “W” corresponding to the four capital movement directions. All other events (such as scrolling, clicking, text selections) were encoded with “X”. Encoding all mouse moves for one SERP interaction pattern such as this resulted in a character sequence. Symbols that occurred contiguously at multiple times were collapsed (e.g., “WEEEWX” →

"WEWX"). The feature "EWEW" encodes whether the user moved the cursor in a right-left-right-left pattern without moving it vertically in between, which is indicative of reading with the mouse (as suggested in Rodden *et al*. [29]). In addition to horizontal reading patterns, we also encoded a number of common sequences (20 in total) that were suggestive of a variety of different SERP engagement strategies (e.g., scrolling vertically with the scroll wheel). Substring searches were performed to determine whether SERP interaction contained the sequence of interest.

4.1.2 Clicks
We computed a range of different features of the clickthrough behavior of users, including the total number of search results that were clicked, the time between the SERP loading and a result click, and the fraction of queries that were abandoned.[2] In a similar way to the featurization of cursor movements, we also computed the total number of hyperlink and non-hyperlink clicks in various AOIs on the SERP, including the number of clicks in the upper and lower search box, the left and right rails, the algorithmic results, and overall across all regions of the SERP.

4.1.3 Scrolling
We also computed features of users' scrolling behavior. These included the total number of scroll events, the number of times they scrolled up, the number of times they scrolled down, the total scroll distance (in pixels), the maximum scroll height (in pixels) referring to the y-coordinate at the top of the viewport relative to the SERP, and the time between SERP load and scroll activity.

4.1.4 Other Features
There were also several other features that were used in this analysis. These include whether the user clicked on the search box (suggesting that they were going to re-query), the number of text selections (total and unique results), and the number of hover previews (total and unique results) requested. Hover previews are a Bing interface feature that provides more information about a search result when requested by a hover over its caption.

Over 80 features are generated for each SERP. We use these features in the analysis presented in the remainder of the paper.

4.2 Individual Differences
To analyze the effects of individual differences we aggregate (average) features per user and cluster users based on those features to identify different patterns of search interaction and groups of users who exhibit those patterns when interacting with SERPs.

Users were identified by an ID stored in a browser cookie. To give us sufficient data from which to base aggregation for each user, we selected all users who had issued at least 20 queries in the time period during which we captured logs. This resulted in a set of 22,084 users whose SERP behavior we analyzed further. Each user in this set issued an average of 39.6 queries (median=31).

4.2.1 Clustering
For each user and for each feature described in Section 4.1, we averaged the feature values across all queries issued by that user in the course of the study. Missing values were properly considered during averaging, e.g., SERPs with no clicks were excluded from the calculation of time-to-first-click.

[2] Note that we had two definitions of SERP abandonment in our analysis: one where there were no clicks anywhere on the page and one where there were no hyperlink clicks. The latter is more traditionally associated with abandonment (e.g., [24]) although we find that the former is more discriminative for clustering.

We used the CLUTO clustering package [20] to identify groups of users who shared similar SERP interaction behaviors. Specifically, we used repeated-bisection clustering with a cosine similarity metric and the ratio of intra- to extra-cluster similarity as the objective function. We found that clusters are fairly stable regardless of the specific clustering or similarity metric. We varied the number of clusters (k) from 2 to 100 and tested within- and between-cluster similarity. We found that the objective function leveled off at k=45, meaning 45 distinct user clusters in our set.

Outlier users were identified and removed by looking for very small clusters (where the number of users was less than ten) with very low extra-cluster similarity at high levels of k. We removed 16 users from the set, leaving us 22,068 users to cluster. To facilitate interpretation of the clusters, we chose a representative set of the 12 most descriptive and discriminative features based on CLUTO output. The following features were selected based on their descriptive value and discriminative power:

- *Time on SERP* (TrailTime): Total time spent on SERP.
- *Clicks*:
  - *HyperlinkResultClickCount*: Number of result clicks.
  - *NonHyperlinkClickCount*: Number of non-link clicks anywhere on the SERP.
  - *TimetoFirstResultClick*: Time between the SERP loading and the first click on a search result.
  - *NoClick*: Whether there was a click (hyperlink or non-hyperlink) on the SERP. We call this abandonment.
- *Re-Query* (ClickInSearchBox): Whether the search box is clicked with the mouse cursor.
- *Scrolling* (Scroll): Whether users scroll (using scrollbar, mouse scroll wheel, or keyboard commands such as Page Down).
- *Cursor*:
  - *FractionTopTenHovered*: Fraction of the top ten result captions (titles/snippets/URLs) that users hover over.
  - *TrailSpeed*: Average speed with which the mouse is moved, in pixels per second.
  - *MedianMouseMovementDistance*: Median distance of individual mouse movements without pauses, in pixels. This helps us understand the degree of focus in the movement. A long distance suggests that the movement is directed.
  - *Reading*: Whether reading pattern is present (Section 4.1.1).
  - *CursorIdle*: Average time the cursor is not moving.

We used these 12 features in a second run of CLUTO, clustering all users based only on this subset. This time, the ratio of within- and between-cluster similarity leveled off at k=6 clusters. We grouped all users in the same cluster together and averaged the feature values. Note that a feature value for a cluster is an average of user averages, thus, every user contributes equally to the cluster average independent of the number of queries they issued.

4.2.2 Cluster Characteristics
Table 1 shows the average values for each feature in each of the six clusters identified. Green (dark) indicates high feature values and yellow (light) low feature values. Labels (e.g., "Long time on SERP") are added to improve interpretability of the clusters. The six clusters are different along a number of dimensions.

Closer inspection of the table reveals three distinct meta-clusters centered on the amount of time and detail with which users spend inspecting the search result page. The three meta-clusters that we identified are: (i) *long (clusters 0 and 1, 11% of users):* careful and detailed SERP examination, many search results hovered on and clicked, lots of scrolling, and signs of reading behavior with the mouse cursor; (ii) *medium (clusters 3 and 4, 15% of users):*

Table 1. Mean average feature values for each of the six user clusters. Green/dark = high values, yellow/light = low values.

	Meta-cluster Name	*Long*		*Medium*		*Short*	
	Cluster	#0	#1	#4	#3	#5	#2
	Number of Users	1178	1283	2240	1167	5645	10555
	Feature						
Time		Long time on SERP				Short time on SERP	
	TrailTime	617.56	342.74	102.83	186.90	55.89	17.51
Clicks			Many R clicks				Few R clicks
	HyperlinkResultClickCount	1.11	1.23	1.17	0.95	0.99	0.88
		Many NH clicks					Few NH clicks
	NonHyperlinkClickCount	0.91	0.92	0.81	0.66	0.58	0.32
			Late first R click				Early first R click
	TimeToFirstResultClick	95.09	325.54	90.91	31.56	38.58	13.52
				Less abandons	More abandons		
	NoClick	0.13	0.12	0.11	0.15	0.14	0.14
Re-Query		More search box					Less search box
	ClickInSearchBox	0.10	0.08	0.07	0.09	0.06	0.04
Scrolling		More scrolling					Less scrolling
	Scroll	0.19	0.21	0.20	0.16	0.14	0.09
Cursor		More hover results					Less hover results
	FractionTopTenHovered	0.30	0.31	0.30	0.28	0.26	0.22
		Slow mouse movements					Fast mouse move
	TrailSpeed	119.07	116.73	127.58	152.52	159.69	230.59
		Short mouse movements					Long mouse move
	MedianMouseMovementDistance	76.91	75.85	77.42	84.05	85.61	90.86
		More reading w/ mouse					Less reading w/ mouse
	Reading	0.06	0.06	0.06	0.06	0.05	0.03
		More cursor idle				Less cursor idle	
	Cursor Idle	54.59	39.85	8.42	18.79	5.36	2.10

intermediate time spent on the SERP and distinguished from other clusters primarily by the amount of abandonment; and *short (clusters 2 and 5, 73% of users):* short time on SERPs, mouse moved quickly and in a focused way, and only a few results inspected.

It is clear that there are differences in how users engage with SERPs. However, it is difficult to isolate user from task differences since most users engage in a variety of tasks during a 13-day time period. Previous work has shown that task and query characteristics impact search behavior. Cole *et al.* [8] found significant differences in how users read results depending on the task. Downey *et al.* [11] showed that user behavior following a query varied significantly with popularity. And, Buscher *et al.* [5] and Cutrell and Guan [9] showed large differences in behavior for navigational vs. informational queries. We now study the impact of task on SERP examination strategies.

4.3 Task Differences

To study task effects, we needed a way to identify different task types at scale. To simplify our analysis we examine individual queries rather than complete search tasks. We focused on navigational and non-navigational queries, which are easy to identify and have been shown to yield different search behaviors in previous work [5][9]. There are other ways to identify queries of different task types, such as if advertisements were shown (commercial queries), the type of inline answer shown (queries with clear intent), etc. We leave such detailed analysis to future work.

To distinguish between navigational and non-navigational queries, we use click entropy [10], which measures the variability in clicked results across users. Click entropy (CE) is calculated as:

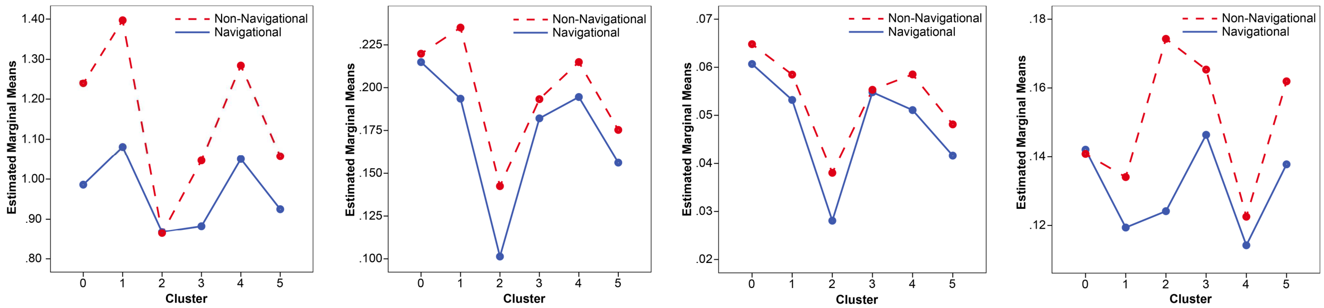
$$CE(q) = -\sum_{URL\ u} p(c_u|q) \cdot log_2(p(c_u|q))$$

where $p(c_u|q)$ is the probability that URL u was clicked following query q. A large click entropy means many pages were clicked for the query, while a small click entropy means only a few were clicked. To divide queries into navigational and non-navigational, we adopted thresholds used by Teevan *et al.* [31] when identifying queries of low and high click entropy: navigational queries < 1.25 click entropy and non-navigational queries > 1.75 click entropy. Click entropy values for all queries in our set were computed across a held out set of one year of Bing query logs. This yielded a set of 514,989 navigational queries and 226,348 non-navigational queries, issued by a total of 22,056 users.

For consistency, we used the same set of 12 features we used in the user clustering analysis and describe how they differ between navigational and non-navigational queries. Table 2 presents the average feature values for each of the two task types.

Table 2. Differences in interaction behavior for different tasks.

Feature		Task Type: *Nav.*	Task Type: *Non-nav.*
TrailTime (secs)		49.59	62.93
Clicks	HyperlinkResultClickCount	0.91	1.02
	NonHyperlinkClickCount	0.49	0.57
	TimetoFirstResultClick (secs)	21.24	29.31
	NoClick	0.13	0.16
ClickInSearchBox		0.11	0.16
Scroll		0.12	0.17
Cursor	FractionTopTenHovered	0.19	0.22
	TrailSpeed (px/second)	180.21	167.43
	MovementDistance (px/move)	80.92	82.94
	Reading	0.04	0.05
	Cursor Idle (secs)	4.31	3.74



(a) HyperlinkResultClickCount *(b) FractionTopTenHovered* *(c) Reading* *(d) NoClick (Abandonment)*

Figure 2. Marginal means for user and task interactions for several features of interest.

Table 2 shows substantial differences in search result page examination behavior between navigational and non-navigational search queries. Given the large sample sizes, all differences across all variables were found to be significant using a multivariate analysis of variance (MANOVA) ($F(12,741314) = 632.54$, $p < .001$). All paired differences for each variable between the task types for each variable were also significant at $p < .002$ using Tukey post-hoc testing. We also computed the effect size of the differences between task types using *partial eta squared* (η_p^2), a commonly used measure of effect size in analyses of variance. The features with the largest effect sizes were the time spent on the SERP (more for non-navigational, η_p^2=.04), the time to first search result click (longer for non-navigational, η_p^2=.02), whether they scrolled on the SERP (more for non-navigational, η_p^2=.01), and the speed with which the mouse was moved (faster for navigational, η_p^2=.01). As expected, users engaged in non-navigational tasks interacted more with the SERP, likely because there may not be a single result to satisfy their needs or they may need to consider many results before selecting a result.

It is clear from the analysis presented in this subsection that there are strong task effects which appear to have a consistent impact on SERP interaction behaviors. So far we have focused on differences in search behavior related to the individual searcher and to the type of search task being attempted. However, there may also be effects attributable to the interaction between users' personal traits and interaction styles and task type which may lead users to examine the SERP differently.

4.4 User × Task Interactions

An important initial step in studying user × task interactions was to compute the fraction of queries for each task type in each of the six user clusters identified in the earlier analysis. Large differences in the distribution of task types found in each user cluster could influence the behavior in each cluster to relate to that task. Analysis of the distribution across navigational and non-navigational queries in each of the user clusters showed that it is approximately similar, with 67–71% of queries labeled navigational in all clusters. This suggests that there is at least a consistent distribution of task types in each user cluster, but says little about the relationship between task and user cluster behavior.

To test for interactions between user and task, we performed a 6 (cluster) × 2 (task type) MANOVA over the 12 dependent variables of interest. The main effects of user and task were both significantly different at $p < .001$ (all $F_{\text{User}}(5,741335) \geq 120.92$, all $F_{\text{Task}}(1,741335) \geq 2.44$). In addition, the user-task interaction was significant for all dependent variables ($F_{\text{User}\times\text{Task}}(60,3706562) = 88.42$, $p < .001$; Tukey post-hoc testing: all $p < .001$). This was expected given the large sample sizes, but many of the interaction effects were small in magnitude. There were some cases where interaction effects size was greater than slight (i.e., $\eta_p^2 \geq 0.04$).

Figure 2 shows examples of the marginal means for four of the dependent variables where there were larger interaction effects between user and task. Non-navigational tasks are represented as red dashed lines and navigational tasks are shown as blue solid lines for each of the six clusters. In each case, most clusters show task-type differences, as often reported in the literature; however one cluster shows little or no task type effects:

1. In Figure 2a, cluster 2 shows no differences in the total number of result clicks per task type, where other clusters of users show a difference of approximately 20%.
2. Figure 2b shows that members of user clusters 1 and 2 explore the results even relatively more deeply for non-navigational queries. Those in user cluster 0 explore search results to the same depth regardless of task type.
3. The reading behavior of members of user cluster 3 is not affected by task type (Figure 2c).
4. The abandonment rate for user cluster 0 is similar regardless of task type and, for cluster 2 abandonment is much higher for non-navigational tasks (Figure 2d).

A better understanding of the nature of the interactions between users and search tasks is important for accurately modeling users and tasks. One way to control for search task is as part of an experiment. Laboratory studies of information seeking behavior (e.g., [5][9]) often control for task type at experiment time. In our case, we must consider the impact of task type retrospectively, by only focusing on user behavior for a particular type of search task. To that end, we now report a new cluster analysis that is restricted to non-navigational tasks to better understand user search patterns for that task type.

4.5 User Clusters for Non-Navigational Tasks

Re-clustering the user data for just non-navigational tasks allowed us to focus our analysis of patterns in the user interaction behavior. We targeted non-navigational tasks rather than navigational since there may be a broader range of information seeking behaviors for those tasks. Given space constraints we cannot share detailed findings for navigational tasks, but as expected there was much more consistency in users' SERP behaviors for those tasks.

We used a similar procedure to that described in Section 4.2.1 to identify the user clusters. However, in this case we extracted only non-navigational queries for each user, and filtered out users who had fewer than 20 non-navigational queries. We clustered these users into k=50 clusters to identify outlier users, i.e., very small clusters with low between cluster similarity. We identified and re-

Table 3. Mean average feature values for each non-navigational user cluster. Green/dark = high values, yellow/light = low values.

	Cluster	*Exhaustive-Passive*	*Exhaustive-Active*	*Economic*
	Number of Users	232	408	1905
	Feature			
Time		Long time on SERP		Short time on SERP
	TrailTime	446.00	156.14	28.03
Clicks		Many R clicks		Few R clicks
	HyperlinkResultClickCount	1.10	1.30	0.85
		Many NH clicks		Few NH clicks
	NonHyperlinkClickCount	0.70	0.85	0.40
			Late first R click	Early first R click
	TimeToFirstResultClick	88.74	144.56	21.05
			Less abandons	More abandons
	NoClick	0.17	0.13	0.20
Re-Query		More search box		Less search box
	ClickInSearchBox	0.08	0.06	0.04
Scrolling		More scrolls		Less scrolls
	Scroll	0.19	0.21	0.13
Cursor		More hover results		Less hover results
	FractionTopTenHovered	0.28	0.29	0.22
		Slow mouse move		Fast mouse move
	TrailSpeed	130.70	118.31	209.11
		Short mouse move		Long mouse move
	MedianMouseMovementDistance	83.75	80.34	94.79
		More reading w/ mouse		Less reading w/ mouse
	Reading	0.05	0.05	0.04
		More cursor idle		Less cursor idle
	Cursor Idle	34.40	9.68	2.21
	Dominant user cluster (% queries from user cluster, meta-cluster)	#0 (30.7%, Lots)	#4 (55.7%, Medium)	#2 (67.8%, Little)

moved one outlier user leaving us with 2,545 users. Next, we reran CLUTO with the same 12 representative features (see above) for the non-navigational tasks for those users, and used the intra- and extra-similarity ratio as the objective function. The ratio between intra- and extra-similarity was maximized at three clusters.

Table 3 shows three distinct emergent clusters of search behavior:

1. *Economic* (75% of users): Users spend little time on the SERP, have focused and fast mouse movements, click quickly, and click on average less than one result per query. Behaviorally, these users are similar to the "economic" users identified in previous work on gaze tracking research [3][12].

2. *Exhaustive-Active* (16% of users): Users who examine the SERP in detail, click a lot (both on hyperlinks and elsewhere), have little cursor idle time, and infrequently abandon. These users are similar to the "exhaustive" users identified previously [3][12].

3. *Exhaustive-Passive* (9% of users): Users who exhibit many of the characteristics as Exhaustive-Active, but spend more time on the SERP, have the cursor idle for a long time, and abandon often.

Interestingly, if we compute the dominant user cluster from which members of each these non-navigational task-related user clusters originated, we see that each cluster in Table 3 corresponds to exactly one of the three meta-clusters in Table 1. The dominant user cluster, the percentage of users from that cluster, and the meta-cluster label are shown in the last row of Table 3. One possible explanation for this finding is that by partitioning by task type, we were able to identify user groupings that were present in Table 1, but were partially hidden due to task effects. In addition to characterizing search behavior using detailed numerical feature values, we also created heat maps of search behavior to determine whether there were any qualitative visual differences in the way that users in each of the clusters inspected the SERPs.

4.5.1 Cluster Heat Maps

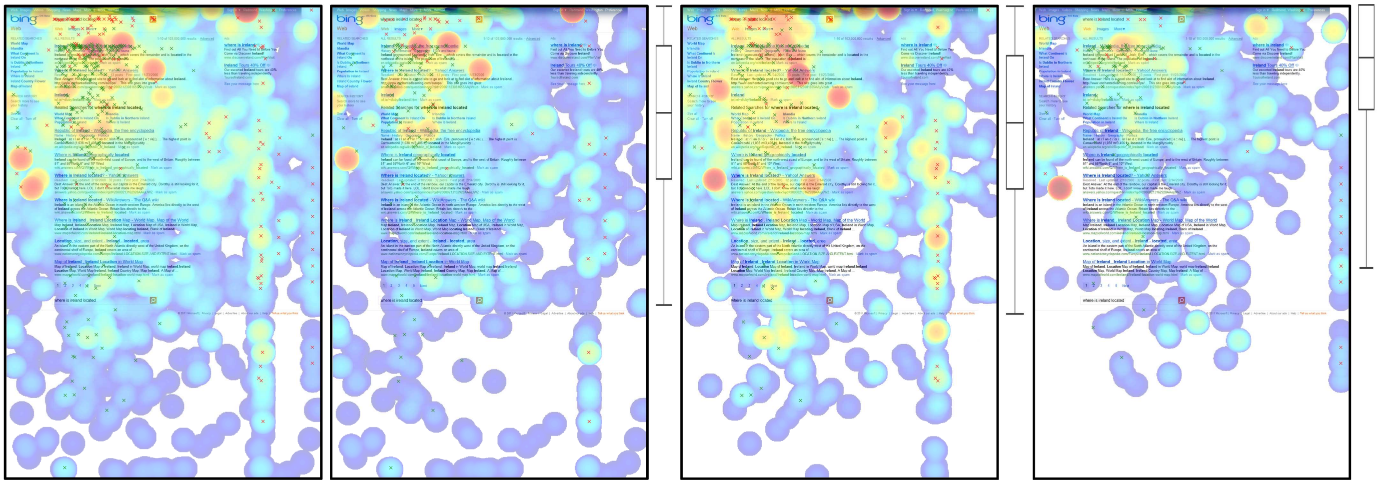
To create these clusters, we randomly selected 100 users from each cluster, and then randomly selected a single query for each of the users. We used these 100 queries to generate the aggregated heat map for each cluster. Figure 3 contains the heat maps for all users (left) and for the three user clusters separately. The spottiness of the heatmap relates to the cursor data sampling rate.

4.5.1.1 Interpreting the Heat Maps

In the heat maps in Figure 3, color represents hover time of the mouse cursor anywhere on the page. For each heatmap this is normalized with respect to the longest existing hover time so that the longest hover time is displayed in dark red. Although small in the figure, clicks are displayed with crosses (×). Green crosses represent hyperlink clicks, red crosses represent non-hyperlink clicks. The image of the SERP in the background is just included as an example for reference. The aggregated impressions come from a large variety of different queries with a variety of SERP layouts, depending on the query. Adjacent to each of the heat maps is a box-and-whisker plot depicting the maximum scroll height reached for each of the clusters. As noted earlier, the scroll position is measured with respect to the uppermost visible pixel in the viewport. Since the average viewport height across all Web browsers in our study was 1142 px (median=743 px), users often had to scroll, but generally only up to one third of the total height of the SERP to see its full contents.

4.5.1.2 Differences in Clusters

The cluster heat maps show fairly consistent differences between the three user clusters that align well with the numeric features reported in Table 3. It is clear from the figures that users in the Economic cluster inspect less of the result page. The deeper examination of Exhaustive-Active and Exhaustive-Passive users is evident in the amount of scrolling that they do, the number of



All *Exhaustive-Active* *Exhaustive-Passive* *Economic*

Figure 3. User cluster heat maps. Box and whisker diagrams show the median, first/third quartiles, and min/max scroll depth.

results hovered over, and the total trail time. The difference between Exhaustive-Active and Exhaustive-Passive users is that Exhaustive-Passive users spend more time over the full result page, as shown by the high *CursorIdle* times in Table 3 and by the more intensely colored heatmap in Figure 3.

Overall, it seems that when we focus on one task type, there are three distinct clusters of search behavior that emerge. These clusters share strong similarities with those identified in previous small-scale gaze tracking studies [3][12]. However, demonstrating the capability to identify similar patterns at scale in more naturalistic cursor tracking logs with more variable tasks is promising.

5. DISCUSSION AND IMPLICATIONS

We have presented a study of individual differences in search result page examination behavior. To our knowledge, this is the first large-scale study of SERP interaction behavior that moves beyond search result click-through. Our findings show that there are cohorts of users who examine search results in a similar way, and that the grouping becomes clearer when we consider task effects. We also showed that there are pronounced task effects that impact how users engage with the SERP and that can interact with users' typical search behaviors. Identifying users with consistent search strategies and patterns is important to understanding how systems are currently being used and create search support.

Our initial analysis revealed six user clusters. However, we also showed that there are strong effects from the type of search task on users' search behavior, as well as strong interaction effects between task and user. When we focused on non-navigational tasks, we found three distinct user clusters who exhibited different result examination behaviors. Promisingly, users exhibited behavioral patterns similar to those found in previous gaze tracking research [3][12], especially the presence of exhaustive and economic groups. Not only do we confirm the existence of these clusters in a naturalistic search setting, but also demonstrate that we can automatically generate them via search engine log analysis.

There are some limitations that we should acknowledge. Since the study was conducted in a naturalistic setting, we do not have control over the search tasks being attempted. Although we automatically labeled task types as navigational or non-navigational, this is a very general task division and it does not incorporate many task nuances that may affect search behavior. The non-navigational tasks in particular are likely to be heterogeneous, and encompass tasks ranging from fact finding to browsing. It would also be possible to identify tasks in other ways based on attributes of the query such as length or popularity, or attributes of the search results such as whether inline answers or advertisements were shown. Developing a finer-grained analysis of task differences is an important direction for future work. We also do not consider the impact of different SERP presentations, as well as users who often issue the same query types (e.g., many queries that return answers, lessening the likelihood that they would click). Future work could address these shortcomings using in-situ methods, where tasks could be assigned and interaction data and user feedback gathered remotely from willing participants. Such methods have been used effectively in previous studies (e.g., [14]).

The cursor-based methods that we have described have the advantage over gaze tracking in that they can be applied on a Web scale, allowing many different types of search behavior to be mined from log data, and significant user cohorts identified. In addition, they provide valuable information about the distribution of visual attention on the SERP that is not available with just hyperlink clicks. This could help improve the search support offered by search engines. For example, we can support the three user groups identified in our analysis in a number of different ways:

Economic users do not spend much time exploring the SERP, have more directed mouse movements, and abandon SERPs often. These users may have clear information needs (that could perhaps be satisfied by a tailored answer on the result page) or are revisiting specific sites. These users could be helped by offering richer answers directly on the result page or direct support for re-finding (as proposed by Teevan *et al.* [32]).

Exhaustive-Active users explore search results in detail and ultimately click on a result. They could benefit from richer summaries that would facilitate decisions about result selection and comparisons between results, as well as an ability to re-rank and explore results based different meta-data (time, topic, author, etc.).

Exhaustive-Passive users explore the results, but are less likely to click on a search result. We could show these users more results or more diverse results to increase the likelihood that they will find something that matches their needs. In addition, we could offer them support for query refinement, since they are also more likely to re-query than other groups.

In addition to supporting users directly, search engines could also use archetypal behaviors for each cohort as additional input to train and evaluate click prediction models (e.g., [7]). Similarly, since cursor tracking provides detailed evidence about the distribution of visual attention to SERP elements, it could be used to evaluate "good abandonment" (where no clicks on the SERP are a good thing [24]), or to measure the impact of new SERP features.

More research is also needed to identify sub-clusters of behavior given more data about each user's SERP interactions, more interaction features for clustering, and more nuanced search task definitions, perhaps spanning multiple queries or sessions.

6. CONCLUSIONS

We have presented the findings of a study on individual and task effects on SERP examination behavior. We analyzed logs containing detailed data on user interactions including clicks, scrolls, and cursor movements for millions of search queries. By clustering the data using these interaction features, we identify individual differences in search behavior, and strong effects of user, task and user-task interaction. When we consider task type (by focusing on non-navigational queries), three distinct user clusters emerge. These clusters share behavioral traits with those identified in laboratory studies, but we observe these without gaze tracking technology and at scale on the Web, opening up a wealth of opportunity for adaptation of the search experience based on individuals' searching behaviors. Future work will expand the feature set and task definitions, explore the use of behavioral patterns to create tailored search experiences, and leverage these rich data for tasks such as click prediction and search result ranking.

REFERENCES

[1] Allen, B. (2000). Individual differences and conundrums of user-centered design. *JASIS*, 51(6): 508–520.

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

[3] Aula, A., Majaranta, P., and Raiha, K.-J. (2005). Eye-tracking reveals personal styles for search result evaluation. *INTERACT*, 1058–1061.

[4] Bhavnani, S. (2001). Important cognitive components of domain-specific search knowledge. *TREC*, 571–578.

[5] Buscher, G., Dumais, S., and Cutrell, E. (2010). The good, the bad and the random: An eye-tracking study of ad quality in Web search. *SIGIR*, 42–49.

[6] Byström, K. and Järvelin, K. (1995). Task complexity affects information seeking and use. *IP&M*, 31(2), 191–213.

[7] Chapelle, O. and Zhang, Y. (2009). A dynamic Bayesian click network click model for Web search ranking. *WWW*, 1–10.

[8] Cole, M., Gwizdka, J., Liu, C., Bierig, R., Belkin, N.J., and Zhang, X. (2011). Task and user effects on reading patterns in information search. *Inter. with Comp.*, 23: 346–362.

[9] Cutrell, E. and Guan, Z. (2007). What are you looking for? An eye-tracking study of information usage in Web search. *CHI*, 407–416.

[10] Dou, Z., Song, R., and Wen, J.R. (2007). A large-scale evaluation and analysis of personalized search strategies. *WWW*, 581–590.

[11] Downey, D., Dumais, S., Liebling, D., and Horvitz, E. (2008). Understanding the relationship between searchers' queries and information goals. *CIKM,* 449–458.

[12] Dumais, S., Buscher, G., and Cutrell, E. (2010). Individual differences in gaze patterns for Web search. *IIiX*, 185–194.

[13] Ford, N., Miller, D., and Moss, N. (2005). Web search strategies and human individual differences. *JASIST,* 56(7): 741–756.

[14] Fox, S., Karnawat, K., Mydland, M., Dumais, S., and White, T. (2005). Evaluating implicit measures to improve the search experience. *ACM TOIS*, 23(2): 147–168.

[15] Guan, Z. and Cutrell, E. (2007). An eye tracking study of the effect of target rank on web search. *CHI*, 417–420.

[16] Guo, Q. and Agichtein, E. (2008). Exploring mouse movements for inferring query intent. *SIGIR*, 707–708.

[17] Guo, Q. and Agichtein, E. (2010). Towards predicting Web searcher gaze position from mouse movements. *Ext. Abstracts CHI*, 3601–3606.

[18] Huang, J., White, R.W., and Dumais, S.T. (2011). No clicks, no problem: Using cursor movements to understand and improve search. *CHI*, 1225–1234.

[19] Joachims, T., Granka, L., Pan, B., Hembrooke, H., and Gay, G. (2005). Accurately interpreting clickthrough data as implicit feedback. *SIGIR*, 154–161.

[20] Karypis, G. Cluto: A clustering toolkit. www.cs.umn.edu/~cluto. Retrieved July 2011.

[21] Kellar, M., Watters, C., and Shepherd, M. (2007). A field study characterizing Web-based information-seeking tasks. *JASIST*, 58(7): 999–1018.

[22] Kohavi, R., Longbotham, R., Sommerfield, D., and Henne, R.M. (2009). Controlled experiments on the Web: Survey and practical guide. *Data Min. Know. Disc.,* 18(1): 140–181.

[23] Lagun, D. and Agichtein, E. (2011). ViewSer: Enabling large-scale remote user studies of Web search examination and interaction. *SIGIR*, 365–374.

[24] Li, J., Huffman, S.B., and Tokuda, A. (2009). Good abandonment in mobile and PC internet search. *SIGIR*, 43–50.

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

[26] Liu., J., Cole., M.J., Liu, C., Bierig, R., Gwizdka, J., Belkin, N.J., Zhang, J., and Zhang, X. (2010). Search behaviors in different task types. *JCDL*, 69–78.

[27] Lorigo, L., Pan, B., Hembrooke, H., Joachims, T., Granka, L., and Gay, G. (2006). The influence of task and gender on search evaluation and behavior using Google. *IP&M,* 42(4): 1123–1131.

[28] Rodden, K. and Fu, X. (2007). Exploring how mouse movements relate to eye movements on web search results pages. *SIGIR Workshop on Web Info. Seek. and Interact.*, 29–32.

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

[30] Saracevic, T. (1991). Individual differences in organizing, searching and retrieving information. *ASIS*, 82–86.

[31] Teevan, J., Dumais, S.T., and Liebling, D. (2008). To personalize or not to personalize: Modeling queries with variation in user intent. *SIGIR*, 163–170.

[32] Teevan, J., Liebling, D.J., and Geetha, G.R. (2011). Understanding and predicting personal navigation. *WSDM*, 85–94.

[33] Thatcher, A. (2008). Web search strategies: The influence of Web experience and task type. *IP&M*, 44(3): 1308–1329.

[34] White, R. and Morris, D. (2007). Investigating the querying and browsing behavior of advanced search engine users. *SIGIR*, 255–262.

[35] White, R. and Drucker, S. (2007). Investigating behavioral variability in Web search. *WWW*, 21–30.

[36] White, R., Dumais, S.T., and Teevan, J. (2009). Characterizing the influence of domains expertise on Web search behavior. *WSDM*, 132–141.