Investigating Behavioral Variability in Web Search

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

Steven M. Drucker
Microsoft Live Labs
One Microsoft Way
Redmond, WA 98052 USA
sdrucker@microsoft.com

ABSTRACT
Understanding the extent to which people's search behaviors differ in terms of the interaction flow and information targeted is important in designing interfaces to help World Wide Web users search more effectively. In this paper we describe a longitudinal log-based study that investigated variability in people's interaction behavior when engaged in search-related activities on the Web. We analyze the search interactions of more than two thousand volunteer users over a five-month period, with the aim of characterizing differences in their interaction styles. The findings of our study suggest that there are dramatic differences in variability in key aspects of the interaction within and between users, and within and between the search queries they submit. Our findings also suggest two classes of extreme user – navigators and explorers – whose search interaction is highly consistent or highly variable. Lessons learned from these users can inform the design of tools to support effective Web-search interactions for everyone.

Categories and Subject Descriptors
H3.3 Information Search and Retrieval: *Search process*

General Terms
Experimentation, Human Factors.

Keywords
Behavioral variability, Web search.

1. INTRODUCTION
Search has emerged as a key enabling technology to facilitate access to information for the general user population of the World Wide Web. Everyday, millions of users submit millions of queries to commercial search engines such as Google, Yahoo!, and Windows Live Search. Such systems generally adopt a "one-size-fits-all" approach to search result presentation, where the same search interface is shown to all users for each query they submit. There is good reason for this: users benefit from familiarity with the interface, and the cost on interface designers is minimized. However, as users perform more tasks using search engines, there is a growing need to understand more precisely what users are doing during the search process. It is only through this understanding that we will be able to build more effective interfaces to cater to more users' queries and searching styles. For example, when we can model and identify consistent behavior, we have a chance to adapt user interfaces to take advantage of predicted behavior. Through the research in areas such as information foraging [27], sensemaking [33], orienteering [25,36] search interface design [8], and information visualization [2], the research community is at the forefront of developing search technology that serves a diverse range of purposes. However, large-scale commercial search engines have not yet been able to effectively apply this rich and varied research, and still favor the traditional ranked-list style of result presentation.

In this paper we present a study of interaction behavior for users engaged in Web search activities that originate with the submission of a query to a search engine. To better understand what users are doing when they are searching, we place a particular emphasis on post-query navigation *trails* (i.e., pages viewed on the click stream following the query being issued). Through client-side logging of 2,527 users over a five-month period, we gathered sufficient interaction log data to perform a detailed analysis of variability in search behavior within and between users and within and between the query statements they issued. Understanding variability (i.e., whether users exhibit consistent interaction patterns in terms of the interaction flow or information targeted) given a user or a query has a range of implications in areas such as the design of search interfaces, predictive document retrieval [19], and user modeling [31]. Although there has been related research on examining user trails [40], studying browsing behavior within Web sites [14], developing user and task models [11], and investigating individual differences in user behavior [10], this is the first study to focus explicitly on behavioral variability in Web search.

In this investigation we wanted to characterize differences in the interaction styles of users, and better understand just how different users' search interactions actually are. In particular, we focus on two research questions: (i) how variable are search interactions within each user and between all users? and (ii) how variable are search interactions within each query and between all queries? To answer these questions we analyze interaction log data for a large sample of users and a set of queries sampled from the logs. In this analysis we focus on interaction patterns, features of the interaction such as time and structure, and features of the information that users interact with, such as the Web domain. As well as providing a better understanding of behavioral variability in Web search, the answers motivate the creation of a tailored set of design recommendations for supporting the most and least variable users and queries that can be offered as optional interface functionality for all Web searchers. The driving force behind this research is a desire to improve the Web-search experience for all users.

In the remainder of this paper we present a discussion of related work, describe the study performed, present the findings and their implications, and conclude.

WWW 2007, May 8–12, 2007, Banff, Alberta, Canada.
ACM 978-1-59593-654-7/07/0005.

2. PREVIOUS WORK

Information-seeking behavior has been studied extensively by the library and information science communities, mainly through the development of models of the search process [1, 21]. In this research, behaviors are often studied as part of a search driven by the desire to complete a task. The task can affect users' ability to formulate queries and interact effectively [4]. Other factors such as domain-specific search knowledge can have a marked effect on search behavior when performing specific tasks such as online shopping and healthcare research [3]. Card *et al.* [5] conducted a user study investigating the effect of information scent (i.e., visual and linguistic cues pertaining to a distal object's information value) on user navigation behavior. They found that there are differences in user behavior for different assigned tasks.

The effect of the individual differences between users has been considered for some time in personalizing search to align better with individual user interests [30]. Research in cognitive and differential psychology [9], user modeling [31], and interface design [23], have all focused on eliciting information about individual users to improve system design. These techniques have generally required access to the users in order to elicit information directly from them via intrusive data collection methods such as questionnaires, interviews, "think alouds", and focus groups. We also examine differences between users, but the approach we adopt is more similar to an observational study.

Observational studies generally involve the experimenter watching the subject perform their normal activities in their natural environment. An advantage of observational studies is that they allow for a deep understanding of naturalistic search behavior. They have been used in a variety of ways from monitoring user interaction with paper documents [20] to users engaged in directed search activities [36]. However, these studies generally focus on small user samples and a limited number of contrived tasks. Our study uses rich interaction log data from a sample size that is several orders of magnitude larger than the typical observational study (allowing trends to be identified and more powerful statistical analysis to be performed), at the expense of personal insight from the subject about their feelings at any given point in time, and an ability to perform moment-by-moment cognition between user clicks [5]. The longitudinal nature of our study allowed us to obtain a large amount of interaction data for each participant, which was particularly useful in analyzing within-user variability.

Other studies have examined Web logs to determine what users are doing during Web search activities. In the search domain this analysis has generally used query logs [15] to gain insight into the types of information people search for and a cursory understanding of how people search. Query logs have been used to address questions such as the repetition of queries to search engines over a period of time [35], and to consider what people find relevant to common queries across individuals [37]. Even when researchers supplement the query log analysis with user surveys [32] these studies are still limited to searches that involve search engines omitting many of post-query activities. Other applications of these logs have been in the prediction patterns of search on temporal patterns of query refinement [17], or the use of these logs in query suggestion [16]. Web site log analysis [6] addresses a broader class of Web search behaviors but conflates undirected browsing behaviors and search.

There has been extensive work on mining and predicting interaction patterns from interaction logs [29] and some research on studying user behavior [34,39]. Part of this research has been the development of a universal law of surfing [14] that suggests user navigation patterns are quite consistent within given Web sites, short surfing sessions are preferred to long ones, and that most users select a small number of hyperlinks in a session. This research has generally focused on user interaction within a single domain, and only limited attention has been given to navigation patterns immediately following the submission of search queries. Research in information foraging theory [27] has gone a long way toward characterizing search behavior, but has mainly focused on the development of cognitive-perceptual simulations of user behavior such as SNIF-ACT [28] to derive a theoretically grounded set of Web site design principles, or the use of information scent to predict surfing patterns of users looking for information on a site [7]. Although there are commonalities with information foraging research, our study specifically addresses behavioral variability not fully handled in the foraging literature.

The very need for users to exhibit more than a trivial number of post-query interactions is linked to the inability of systems to fully understand the information needs of their users. As has been suggested already [36] the "perfect" search engine (i.e., a search engine that returns exactly what is sought given a fully specified information need) may address some problems, but there may be circumstances where users are unable to specify their information needs at a level to make systems effective. Instead, it has been observed that users exhibit a style of interaction known as orienteering [25] whereby they use the search engine to parachute them into a relevant part of the information space, then depend on their recall and recognition skills to help them from there. Many of the trails we use in this study exhibit traits of orienteering behavior. Features of these trails such as the time taken to traverse them can provide insight into patterns of interaction that can be useful in designing search systems.

Previous work in this area is rich and varied. However, our study is unique in that we focus solely on search behavior, do so over a five-month period (longer than any other study of this nature), and target behavioral variability with the aim of better understanding the degree of difference in the interactions for different users and different queries. Evidence of significant behavioral variability would make a case for tailored support for users and queries, and consistency would make a case for generic support across all users and queries, similar to what is offered by commercial search engines presently. We now describe our methodology.

3. METHODS

We conducted a log-based study of variability in post-query interaction when engaged in Web search activities on commercial search engines. In this section we describe the study in more depth, beginning with details of participation.

3.1 Participation

When downloading a partner client-side application from the Web, users were asked to consent to their within-browser interactions being logged and later analyzed[1]; 3,291 people agreed. We logged their interaction over a period of five months from December 2005 to April 2006. We chose to elicit no personal information from participants and we were not able to communicate with them directly. The reasons for this were two-fold: to safeguard privacy we did not want to record any more personal information about participants than was necessary to perform the analysis planned before the experiment, and we did not want to unduly influence

[1] It is important to note that if subjects did not provide their consent, their interaction was not logged.

participants' search behavior by making them conscious that their interaction may be tied to personal information about them (even though this was not the case). To maintain the realism that is an advantage of a log-based study, it was important we minimize such potential biasing effects. To this end, we did not elicit additional information from participants via questionnaires or interviews.

The logging mechanism was deployed as a plugin to the Microsoft Internet Explorer browser, and wrote an entry to a remote server each time a new Web page was loaded in the browser. The information contained in these log entries included a unique (anonymous) identifier for the participant, a timestamp for each page they viewed (in the participant's time zone), a unique browser window identifier (to resolve ambiguities when determining the browser in which a page was viewed), and the URL of the Web page visited. For privacy reasons, we did not log the contents of the pages being viewed or any pages viewed over a secure connection. The user trace information gathered was sufficient to reconstruct the trails followed by users at a page-level granularity (i.e., one log entry with unique participant identifier, time, etc. per page view). Over the five-month duration of the study, the 3,291 participants viewed millions of Web pages, submitted millions of queries, and spent many hours online. However, further analysis of the logs revealed that whilst participants were generally fairly active during this time, there were a significant proportion of participants (23%) who engaged in less than 50 search sessions during the five-month period. Many of these participants disappeared from our logs after one or two months, perhaps because they downloaded a new browser, installed an additional plugin that may have interfered with the logging, stopped searching, or opted-out by uninstalling the plug-in. Since we required many search sessions from each participant to study levels of variability in interaction patterns, we removed those participants from the pool, leaving 2,527 participants whose behaviors we examined in more detail.

3.2 Search Trail Extraction

Within each participant, interaction logs were grouped based on browser identifier information. Within each browser instance, participant navigation was depicted in the form of a continuous path we refer to as a *browser trail*, from the first to the last Web page visited in that browser. Located within some of these browser trails were *search trails* that originated with the submission of a query to a major commercial search engine such as Google, Yahoo!, MSN Search, and Ask.[2] It is these search trails extracted from participants' interaction logs that we use in this study.

Search trails originate with a directed search (i.e., a query issued to a search engine), and proceed until a point of termination where it is assumed that the user has completed their information-seeking activity. Trails can contain multiple query iterations, and must contain pages that are either: search result pages, visits to search engine homepages, or connected to a search result page via a hyperlink trail. To reduce the amount of "noise" from pages unrelated to the active search task (that may pollute our data) we introduced some termination activities that we used to determine the end-points of search trails:

- *Return to homepage:* Returning to a homepage was assumed to mark the termination of a search trail. Although we had no access to participants' browser settings, analysis of whether the page appeared consistently first when a new browser window was opened, allowed us to approximate each participant's homepage.
- *Check email or logon to service:* Checking Web-based e-mail, or logging-in to online services such as "myspace.com" or "del.ico.us", was used as an indicator that the search trail had terminated.
- *Type URL or visit bookmarked pages:* Entering a URL directly into the address bar of the browser, or selecting a bookmark, terminated the search trail. The only exceptions were visits to search engine homepages (e.g., http://www.google.com), which may be a necessary part of the current search activity, particularly if participants decide to change search engine mid-trail.
- *Page timeout:* If the display time for any page exceeded 30 minutes this was assumed to mark the termination of a search trail. Timeouts of this duration have been used to demarcate sessions in previous research [6].
- *Close browser window*

These trail termination points are somewhat heuristic, and some may be related to the active search task, e.g., checking email to support task resolution, or running multiple searches on the same topic concurrently in different browser windows (or different tabs within the same window). However, we felt that removing unrelated noise from the trails outweighs the cost of potentially truncating some trails early.

To demonstrate how search trails are constructed, we present an example of how the trail would be extracted given a candidate browser trail. To simplify the exposition, we express the browser trail as a Web behavior graph [5], shown in Figure 1.[3] This graph represents user activity within a browser trail, from their homepage (*S1*) through to the point at which they close the browser (*X*). The nodes of the graph represent Web pages that the user has visited: rectangles represent page views and rounded rectangles represent search engine result pages. Vertical lines represent backtracking to an earlier state (such as returning to a page of results in a search engine after following an unproductive link). A "back" arrow, such as that below *S5* implies that the user is about to revisit a page seen earlier in the browser trail. Time runs left to right and then from top to bottom. The region of the graph shown in gray represents a Web-based email service, in this case Microsoft's "hotmail.com".

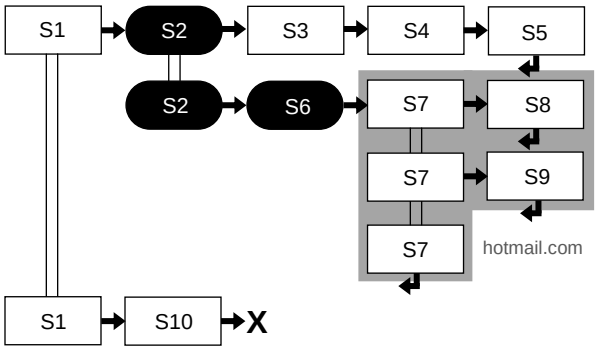

Figure 1. Browser trail as Web behavior graph.

[2] Although subjects may have conducted other directed search activities within Web sites (through a "site search" option) we were unable to consistently log those events.

[3] Web behavior graphs are a variant of problem behavior graphs [24], and are useful for viewing navigation patterns.

In the example browser trail shown above, the user is pursuing information related to their original search query. As they navigate, they check their Web-based email (shown in gray), return to their homepage (*S1*), viewed one page linked from that page, and close the browser window (*X*). Given this browser trail, the search trail would run from *S2* (the submission of the first query) to *S6* (the last page viewed before email checking). The visit to the Web-based email service matches one of the five termination criteria described earlier in this section. The search trail in the example is therefore *S2→S3→S4→S5→S2→S6*.

Terminating the trail on an event such as checking email or returning to a homepage was preferred to trails delimited on each query submission. Since searches generally involve multiple query iterations, running trails over multiple iterations allows us to analyze richer interaction patterns than for individual queries. In addition, it is beyond the scope of this work to try to determine how to group or separate successive queries. Given the nature of the interaction logs generated by our client-side application, we were able to extract search trails relatively easily using the approach described here. As such we circumvented the need to do this with probabilistic models of behavior [e.g., 29].

Extracting search trails using this approach also goes some way toward handling multi-tasking, where users run multiple searches concurrently. Our search trail extraction is based on first extracting a browser trail for each instance of the Web browser, then looking within that trail for the first query, and extracting the search trail from that point onwards. Since users engaged in multi-tasking operations generally open a new browser window (or tab) for each task, most tasks have their own browser trail, and through our methodology, in turn a separate search trail.

3.3 Query Selection

During the study, participants issued around three million query statements, and followed approximately a half million search trails where at least one search result was clicked.[4] 15.6% of query statements appeared only once across all participants in the five-month period (i.e., were singletons). Although the presence of a Zipfian distribution is common in Web search query logs, its reoccurrence in the logs used for this analysis was potentially problematic. Any reasonable statistical analysis required multiple instances of each query, and preferably multiple occurrences of each query statement from multiple participants. From the set of unique *initial*[5] query statements, we selected a subset of query statements that were issued at least 15 times and by at least 15 unique participants.[6] These thresholds gave us sufficient data and overlap between participants and queries. The resultant set of queries represents 10.2% of all initial query statements, and contains the 385 most popular query statements (each potentially submitted by many users, many times). Ideally we would have liked to use all queries in our analysis. However, given the large number of queries appearing only a few times, this was not possible. Although the 385 queries were the most popular, in terms of the proportion of navigational, informational, and resource queries [32] their distribution was equivalent to query logs many orders of magnitude larger. We now describe the findings.

[4] On average there were 5 queries per search trail.

[5] Note that the "initial" query is submitted to initiate a search trail (i.e., the first query iteration).

[6] The value 15 was selected as it seemed to represent a natural threshold in the query frequencies, and gave us a reasonable number of instances per query.

4. FINDINGS

We first present an overview of interaction behavior for all participants over all queries for the five-month period. We then analyze the variance in search behavior for participants and queries. Results of statistical analysis are presented at a .05 level of significance unless otherwise stated. Parametric statistical testing is used where appropriate.

4.1 Overview of Search Behavior

Over the five month period the 2,527 participants viewed approximately 80 million Web pages, 12.5% of which were part of a search trail as defined earlier. The remainder of time was spent checking electronic mail, conducting e-commerce, and other browsing events such as following hyperlinks from their homepage. Based on the URL, we automatically classified pages that lie on the search trails into two types: *search* (*S*) (a directed search involving a recognized commercial search engine) and *browse* (*B*) (a view of a page that lies somewhere on the click path flowing away from a search results page). Between these page types there was also two types of transitions: *forward* (*f*) (where a user clicks a hyperlink to visit a page not previously visited on the search trail, e.g., the move from *S4→S5* in Figure 1), and *backward* (*b*) (where a user revisits a page on the trail, e.g., the move from *S5→S2* in Figure 1). Figure 2 shows the proportion of interaction events that fall into each of four categories: *forward-to-search*, *backward-to-search*, *forward-to-browse*, and *backward-to-browse*. For example, the chart shows that the least popular interaction was returning to a previous search page (only 8% of interactions involved this operation).

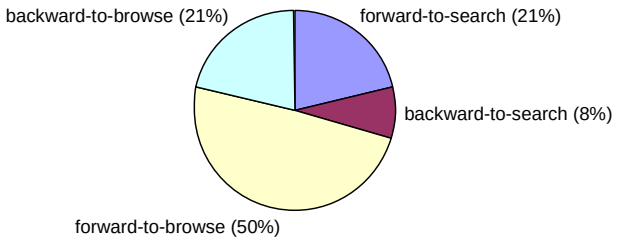

Figure 2. Percentage of interactions classified as searching and browsing.

Less than one third (29%) of interaction is with search result pages, and the remainder is with pages that lie on the hyperlink trail from a search result page. These findings not only stress the importance of post-query interaction in search (as also suggested in [36]), but also demonstrates the volume of interaction behavior that is available for analysis when both navigation *and* querying behaviors are considered.

We now focus on the variability in interaction behaviors for queries and users. For this, and all subsequent analyses, we group the data in two ways: *by user* (i.e., all search trails for each participant), and *by query* (i.e., all search trails for each query). M and SD are used to denote the mean average and standard deviation respectively.

4.2 User Variability

We were interested in investigating whether user interaction was consistent within each user and between all users, regardless of the queries they submitted. This is an important question with possible implications in areas such as personalization and predictive information retrieval. For example, should user interaction be fairly consistent within each user, but variable across users, this strengthens the case for personalized search [30]. We divided the

analysis into three parts in which we first studied differences in interaction patterns, then the inclusion of additional trail features, and finally the variation in the domains participants visited.

4.2.1 Differences in Interaction Patterns

To characterize the search trails we needed a way to represent user queries that would allow comparisons of interaction *behaviors* to be conducted. We modified a method used in related work [6], that represented trails as strings, and then compute the Levenshtein Distance (LD) [18] between trails represented in this way. LD is a method for judging the closeness of two arbitrary length strings based upon the number of insertions, deletions, and substitutions necessary to convert one string to another. This provides a way to estimate variance in interaction patterns for each user and query.

Since the space of possible pages visited is potentially large for many participants, it is simply not practical to give each unique page its own unique symbol. Instead, we represent the pages viewed on the trails based on their type (i.e., either search (*S*) or browse (*B*)) and the transition between them (i.e., either forward (implied by the ordering)[7] or back (*b*)). For example, the trail illustrated by the graph in Figure 1 is represented by the string "*BSBBBbSSBBbBBbBbBB*".

To approximate within-user *interaction variance* we computed LD from each search trail followed by that participant to every other search trail followed by that participant. We calculated the average distance across all trails a participant follows, and assumed that the trail with the smallest average distance from all trails was most representative of the interaction patterns of that participant. The average distance computed in this calculation is used as a measure of interaction variance. That is, if the most representative trail is a low average distance from all other trails for that user then it is assumed that there is low variance in the search interaction patterns of that particular user. In contrast, if the average distance from the representative is high, then it is assumed there is high variance in the search related interaction behavior of that user. To clarify this procedure we present an example.

Example: Given the following three search trails, we use the approach described in this section to determine the most representative trail (and its variance).

Step 1: *Represent trails as strings.*

1: S1→S2→S3→S2→S5→S6 = *SSBbSBS*
2: S1'→S2'→S3'→S2'→S5'→S1'→S6' = *SBBbBSbSS*
3: S1''→S2''→S3''→S4''→S5'' = *SBBBB*

Step 2: *Calculate average distance between strings.*

From Trail 1:	From Trail 2:	From Trail 3:
LD (1,2) = 4	LD (2,1) = 4	LD (3,1) = 4
LD (1,3) = 4	LD (2,3) = 5	LD (3,2) = 5
Average = 4	Average = 4.5	Average = 4.5

Step 3: *Select most representative trail.*

Trail 1 has minimum average distance and variance = 4

---

[7] Since our earlier analysis showed 70% of search interaction events were a forward motion, representing this explicitly in the notation would add redundancy with no effect on the LD.

The use of this measure as an approximation in the level of variance in interaction behavior allowed us to compute an interaction variance value for each of the 2,527 participants across all of their search trails. The average interaction variance was 20.1 (SD = 11.8, Max = 94.4, Min = 3.2, Median = 16). Exploratory data analysis revealed that the data were not normally distributed, and exhibited a positive skew whereby many participants had a low interaction variance (e.g., 12% of participants had an interaction variance less than 10, but only 2% of participants had a variance of more than 80).[8] This suggests that most participants interacted fairly consistently, and that across all queries they issued, most users' search interaction style did not differ greatly.

We were particularly interested in whether there was any relationship between interaction variance and features of the search trails that we felt were typical of exploratory search activity, namely the branchiness (i.e., the number of re-visits to previous pages that were then followed by a forward motion), and number of query iterations. A multiple regression performed between these three features revealed a strong relationship ($R^2 = .32$, $F(2, 2524) = .594$, $p \leq .001$). This suggests that interaction variance is a good predictor of aspects of the search process that go beyond the patterns of interaction represented in textual strings (e.g., such as querying behavior). This gives support to the use of our approach based on LD to classify search trails. The regression also highlighted participants whose interaction variance was highly consistent (i.e., the LD was very low) and those where it was highly variable (i.e., the LD was very high). Participants that lay at these extremes were classified as "navigators" and "explorers". We now describe these two classes in more detail.

Navigators (low variance): These users have consistent interaction patterns in the trails they follow. That is, many of their search trails look similar when they are reduced to the representation used to compute LD. Users whose interactions were consistent seemingly interacted in a particular way. Further analysis revealed three additional attributes of navigators' search trails: (i) they exhibited few deviations or regressions, (ii) they appeared to tackle problems sequentially, and (iii) they were more likely to revisit domains. We name these extremely consistent users "navigators" since it appears they follow a seemingly direct path from query submission to problem resolution. In Figure 3 we use the Web behavior graphs introduced earlier and present a typical search trail for navigators. The shaded regions represent Web domains, the black rounded nodes represent search engine result pages, the double vertical lines (i.e., | |) represent revisits to pages previously encountered on the search trail, and the labels on each node represent the step number.

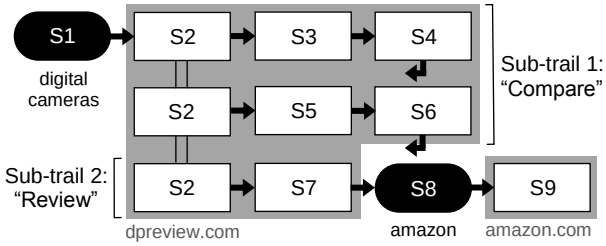

Figure 3. Typical search trails followed by "navigators".

In the example, the user appears to want to select and purchase a digital camera. They complete two sub-tasks – compare cameras and read reviews of a particular camera – within the first domain

[8] A log-transform was used correct the data to normal form.

"dpreview.com", a popular digital photography review site, before proceeding to issue a new query and then browse to the second domain, "amazon.com", perhaps to purchase the item. In this example, the Web page at *S2* seems to be a particularly important interaction hub within "dpreview.com". Branching points such as these appear to be important to support the "building block" strategy evident in most of the searches that navigators conducted.

It is worth noting that we would expect most users to exhibit navigator-style behaviors when they attempt a well-defined fact-finding task. However, navigators represent an extreme case of users since almost all of their search interactions are this way, regardless of the query and even though there were no notable differences in the types of queries submitted by navigators compared to all other users.

Explorers (high variance): These users have variable interaction patterns in the trails they follow. That is many of the search trails for each of these users looked different when they were reduced to the representation used to compute LD. Further analysis revealed three additional attributes of explorers' search trails: (i) they tended to branch frequently, (ii) they submitted many queries during a search session, and (iii) they visit many new domains. We name these extremely inconsistent users "explorers" since they appear to utilize multiple strategies concurrently when searching for information, and do not follow a direct path from problem specification to resolution. In Figure 4 we show an example of a search trail that is typical of an explorer.

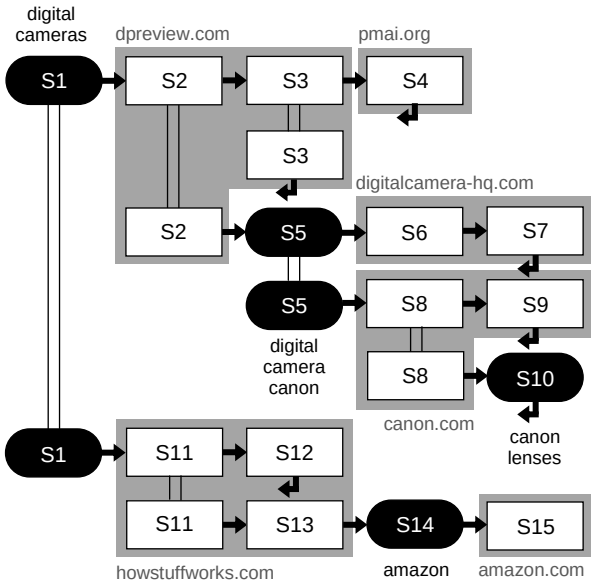

Figure 4. Typical search trail followed by an "explorer".

As can be seen in the figure, the explorer visits multiple domains and submits many queries during the course of their search. In this case, this includes a brief visit to the Web site of the Photo Marketing Association International (pmai.org). This behavior should be contrasted with that of the navigator in Figure 3. Both trails start with the same query and end at the same domain (i.e., "amazon.com"), but their interaction in-between is much different. Explorers jumped between different domains frequently, and seem to be targeting multiple aspects of the search task simultaneously.

Once again, it is worth noting that we would expect these behaviors to be exhibited by all users depending on the query. For example, in complex sensemaking tasks an exploration strategy such as shown in Figure 4 may be appropriate. However, explorers represent an extreme case since almost all of their search interactions are this way, regardless of the queries they submit and even though there were no notable differences in the distribution of queries issued by explorers compared to all other users. Explorers may be more likely to be distracted by interesting links and serendipitous information encounters, perhaps even in the form of contextual advertising. There may be more variance in their search trails simply because the trails they follow are significantly more complex than those followed by the general user population.

Following the application of a logarithmic transform to make the data normally distributed, we devised thresholds that corresponded to the 95% confidence interval to give us upper and lower bounds on interaction variance (translated back into the original LD) of 14 and 75 respectively. These provide an approximation of whether the interaction variance of a participant is significantly different (at $\alpha = .05$) from the mean average variance of all users. According to this classification, approximately 17% of participants had an interaction variance of 14 or less (meaning they were classified as "navigators"), and approximately 3% had an interaction variance of 75 or above (meaning they were classified as "explorers"). The remaining 80% of participants lay somewhere between navigators and explorers (although the positive skew suggests that most were similar to navigators), and further classification of them may be possible with more detailed analysis planned for future work.

Navigators and explorers represent two extremes of interaction variance. However, improving our understanding of the behavior at these extremes can teach us to build more effective search solutions for them that are transferable to less extreme users and situations.

4.2.2 Differences in Trail Features

Until now we have focused solely on patterns of interaction and ignored additional features of the search trails that may be useful in characterizing interaction variance. To facilitate more complete analysis we extracted the following six observable features for the each search trail:

- *Time:* Amount of time spent (in seconds) on a trail.
- *Number of queries:* The number of queries that were submitted during a trail.
- *Number of steps:* The number of pages viewed in a trail, including all searches and revisits.
- *Number of revisits:* The number of revisits to a page viewed earlier in the trail. Revisits to pages viewed previously in other trails are disregarded.
- *Number of branches:* The number of times a subject revisited a previous page on the trail and then proceeds with forward motion to view another page; this is subtly different from the number of revisits. To qualify as a "branch", the user must navigate to a page following the back operation and prior to the next back operation (if any). For example, the browser trail illustrated by the Web behavior graph in Figure 1 has four branches yet five rows in the graph. Row 4 does not constitute a branch since there is no forward motion.
- *Average branch length:* The average number of steps in each branch in the trail.

These features can give insight into aspects of search behavior not apparent from the interaction patterns used in the previous section. In Table 1 we present summary statistics for the trail features, averaged across all users and all queries. The values shown can be useful for investigating differences between these two groupings.

Table 1. Descriptive statistics on search trail features.

Search trail feature	User		Query	
	M	SD	M	SD
Time	476.4	214.4	435.4	192.3
Number of queries	4.3	3.2	4.0	3.0
Number of steps	17.7	8.5	16.5	6.2
Number of revisits	5.1	4.2	4.3	2.1
Number of branches	4.1	2.9	3.6	1.5
Average branch length	4.2	0.7	4.5	0.6

Across all users and all queries the average values for each of the features shown in the table are similar, although slightly lower for queries than for users. Variance in the query grouping is generally lower than the user grouping for features related to the length of the trails and their branchiness. We used a factor analysis to determine whether these seven possible features (the six features mentioned above plus the LD) could be explained with less than seven factors. Factor analysis is a statistical technique used to explain variability among observed random variables in terms of fewer unobserved random variables called factors. We employed a variant of factor analysis, known as exploratory factor analysis, which assumes *a priori* that any measure may be associated with any factor. Generally, the fewer factors identified, the less the data varies.

To perform this analysis, we computed the intercorrelations between the seven features. Our findings showed that all features were correlated at a statistically significant level (all $\underline{t}$ (2525) $\geq$ 4.49, all $\underline{p} \leq .001$)[9] with the exception of *average branch length* and *number of revisits*, which exhibited no correlation between them. The factor analysis revealed the presence of three factors that could account for 80.6% of the variance between users:

- *Forward and backward motion:* The most powerful factor, contributing 52.5% of the variance. It appears to represent a very basic dimension that relates to clickthrough and "back" operations.
- *Branchiness:* This accounts for 17.4% of the variance and represents the extent to which the user follows sub-trails within the search trails. More sub-trails generally imply the existence of more facets in the search task.
- *Time:* This accounts for 10.7% of the variance and represents the amount of time taken to traverse the search trail.

These factors appear important in characterizing the interaction variance between users, and may be useful in building probabilistic models of user behavior that help to automatically determine how variable a user's interaction is compared to the general population.

As well as considering the nature of the search interaction it is also important to consider which domains participants were actually visiting during their searches. This expands our definition of variance beyond observable aspects of the interaction to include an important dimension focused on what users were interacting with. In the next section we analyze the domains that users visited.

[9] In addition, Cronbach's alpha = .811.

4.2.3 Differences in Domains Visited

Participants who returned to the same domain many times (perhaps even following the same trail to get there) may exhibit less variance in their interaction patterns than those who revisited previous trails fewer times. This may be simply because of improved knowledge of what search results to expect, and improved knowledge of domains they are exploring. In this part of the analysis we are not concerned with the order in which participants viewed pages or the structure of their interaction, rather the extent to which the pages which participants interacted with were unique. For each of the 2,527 participants, across all queries they issued, we extracted the names of domains they had visited in the duration of this study. We then computed *domain variance* as a function of the number of unique domains visited divided by the total number of domains visited. The value that results from this computation lies between 0 and 1 (inclusive), with a lower value indicating more consistency in the domains they visited.

The results show that participants generally viewed a diverse range of Web pages during the study ($\underline{M}$ = .29, $\underline{SD}$ = .21). Although the domain variance for most participants was grouped around the mean, 2% of participants had a variance of .9 or more, suggesting that 90% the domains they visited were unique. At the opposite end of the spectrum, 17% of all participants had a variance of .1 or less, suggesting many of the domains they visited were re-visits. There was a strong positive correlation between domain variance and interaction variance between participants ($\underline{r}$ (2525) = .42, $\underline{p}$ < .001). It is important to note that those participants we had already classified as explorers appeared also more likely to visit different domains during their searches. In contrast, those we had already classified as navigators were more likely to visit the same domains.

User interaction appears fairly consistent within and between users in terms of their interaction and the domains they visit. Navigators and explorers represent the extremes of interaction variance, but we can learn from them to support all users when they encounter tasks that may require them to act like a navigator or explorer. We now perform a similar analysis for queries rather than users.

4.3 Query Variability

In the second part of the analysis we now consider interaction variance within and between the 385 test queries sampled from our participants' logs. As stated earlier, these test queries comprised 10.2% of all initial queries submitted by subjects to start search trails. If interaction across queries was variable it may point to the development of tailored search solutions for queries within which interaction is highly consistent or highly variable. Examples of this support could be the provision of guided tours [38], the use of teleportation to facilitate rapid access to frequently visited destinations [36], or query segmentation for tailored search engine training or ranking operations.

4.3.1 Differences in Interaction Patterns

In a similar way to the analysis performed earlier we computed the LD from each search trail to each other possible trail for each test query, computed the average distance, and took the trail with smallest distance as the best example trail for that query, and the minimum distance from that trail to every other trail as an approximate measure of interaction variance. The results of this analysis yielded a variance value for each of the test queries across all search trails for which the query occurred in the first directed search. The average variance was 18.7 ($\underline{SD}$ = 9.1), the maximum interaction variance was 98.5 (for the query "sexy"), and the minimum interaction variance of was 2.4 (for the query "www.msn.com"). Further inspection of the queries revealed that

those with the smallest amounts of variance in their interaction behaviors were generally navigational and served to get people to a particular Web site. In contrast, those with the highest variance were generally undirected exploratory queries to obtain general topic knowledge, and queries where peoples' tastes may differ (e.g., pornography, travel, art). Of the 20 terms with lowest variance, 18 were *navigational* (i.e., the immediate intent is to visit a particular Web site), and two were *informational* (i.e., the intent was to acquire information from one or more Web pages). In contrast, of the 20 terms with highest variance, 16 were informational, and only four were navigational.[10] It appears that variation in search behavior is certainly lower for queries with a known destination. An additional application of the LD measure may be in the development of an automated query classification mechanism, to segment queries based on post-query browsing behavior.

4.3.2 Differences in Trail Features

Across all queries we computed the intercorrelations between the seven features used in Section 4.2.2, and applied a factor analysis. All features were correlated at a statistically significant level (all $t(383) \geq 2.72$, all $p \leq .007$)[11] with the exception of *time* and *number of query iterations*, which exhibited no correlation between them. Time taken to follow a search trail appears unrelated to the number of queries submitted on the trail. This may be because query operations form only 29% of all trail steps; the majority of the interaction (and the time spent) is with pages beyond search result pages. The factor analysis revealed the presence of two factors that could account for 78.4% of the variance. To reliably represent the variance in attributes of participant behavior in search trails between queries we only need only use "*forward and backward motion*" (58.5% of the variance) and "*time*" (19.9% of the variance). These two factors appear important in characterizing the variance in the interaction between the test queries.

4.3.3 Differences in Domains Visited

We conducted an analysis of the variance in the domains visited for all 385 test queries in a similar way to that conducted in Section 4.2.3. The average domain variance per query was .14 ($SD = .08$). A high domain variance implies that participants had to visit a diverse range of domains to find the information they were looking for. The most variable queries were generally broad informational queries, such as "chat" and "search", whereas the least variable queries were generally navigational and served to get users to a particular Web site. The results of additional analysis suggested that although many of the queries with the lowest domain variance were navigational and those with highest domain variance were informational, there was a negative correlation between the domain and interaction variance ($r(383) = -.31$, $p < .001$). The queries for which interaction varied most were generally those with less variance in the domains visited. For these queries users may be interacting extensively with a few popular domains that contain many Web pages and have large numbers of hyperlinks between their constituent pages. Examples of such sites include "msn.com", "amazon.com", "youtube.com", and "yahoo.com".

[10] Prior to analyzing the data, we classified all 385 queries in the test set into navigational, informational, or resource using the classification outlined by Rose and Levinson [32]. The queries were distributed as 33% navigational, 61% informational, and 6% resource (e.g., find sites offering a service). This distribution is representative of query distributions in much larger query logs [32].

[11] In addition, Cronbach's alpha = .856.

Variance in user interactions and domains visited appears related to the nature of the search query. Enhanced versions of the variance measures we used may offer a viable alternative to labor-intensive process of manually classifying queries. We now discuss how the findings of our study relate to other studies of information-seeking behavior, and provide implications for search interface design.

4. DISCUSSION AND IMPLICATIONS

Our study attempted to characterize differences in the interaction patterns and information targets attributable to users and queries. The findings emphasized the importance of research such as this and the importance of treating some users and queries differently. We begin this section by relating our findings to previous work.

Ford *et al.* [12] conducted a study of 111 peoples' search behaviors, and showed that cognitive styles (i.e., tendencies displayed by individuals consistently to adopt a particular type of information processing strategy) appeared to influence search behavior. Through user experimentation, Pask [26] identified two cognitive styles – *holist* and *serialist* – that can be used to classify individuals based on their learning strategy. Holists are cognitively complex, tend to exhibit a global approach to learning, and elect to investigate relationships between multiple objects during the learning process. Given the variability in their interaction patterns, the branchy nature of their trails, and multiple query statements they submit, it is conceivable that our "explorers" may share some attributes with holists. In support of this claim, Ford *et al*. [12] found that holists appeared more engaged in exploratory searches and valued serendipitous information encounters. These are traits that we imagine our explorers may share. In contrast, serialists tend to use a local learning strategy, examining one thing at a time, and concentrating on separate topics and the logical sequences linking them. They may be similar in some respects to our "navigators", given that their interaction patterns were generally consistent, and they exhibited few deviations or regressions from a direct route to their information target. However, cognitive styles may not be the only contributor to differences in interaction behavior.

Hölscher and Strube [13] conducted a study of search behaviors, and found that people's background, knowledge of the Web, and search experience can greatly influence their search behavior. In addition, the Teevan *et al.* [36] study described earlier in this paper, classified the 13 participants as "filers" and "pilers" [20] based on the information organization strategy they employ. Filers tend to use a rigid structure and pilers tend to maintain information in an unstructured way. They suggested that filers and pilers tended to rely on different search tactics in terms of the frequency of searching and the number of keyword searches performed. Navigators appear to be exhibiting a style of interaction more typical of a structured information organization strategy, with focused directed searches and topical coherence in the search trails. In contrast, explorers' tactics appear more suited to an unstructured information organization strategy, with a high interaction variance and (re-)visits to multiple domains.

It is likely the case that cognitive styles, search task, background, information organization strategies, and search experience, to name but a few, are behind the variations in behavior we witnessed in our study. The studies mentioned above have benefited from using both qualitative and quantitative methods to analyze data gathered. While our study is tempered by lack of such information about our participants, our study is several orders of magnitude larger than those described above, allowing trends in the interaction behaviors to be more easily identified.

The current "one-size-fits-all" approach to search interface design supports the users and queries in many cases. However, much can be learned from extreme users and extreme queries that can be used to supplement this generic approach. From our analysis, we determined that extreme users (i.e., those whose interaction behavior was extremely consistent or variable) represented approximately 20% of our participant sample. Extreme queries (i.e., those that promote interaction that is extremely consistent or variable) represented 18% of our test set. These cases appear to represent a significant proportion of the user and query population that can assist in developing tailored interfaces for the benefit of the masses. Rather than just developing interfaces for the extreme 20%, we aim to use the characteristics of the interaction of these users and for these queries to identify what support to offer. The outcome is a set of recommendations tailored to the extremes, but if implemented as an optional part of existing search systems, could support the 80% of users who exhibit extreme behavior infrequently. In the remainder of this section we discuss the implications of our research in two parts: (i) how could we identify (on a large scale) the level of variance in a user or query? and (ii) what is the appropriate system support to offer in each case?

Supporting users based on their individual information-seeking strategies (and variance between these strategies) requires detailed information about their post-query behavior over a period of time. On a large-scale it is impractical to have users personally describe their interaction behavior across all searches they conduct. Even if we could elicit this information from them, it is unlikely that we will get a true indication of their behavior. The use of log-based approaches, such as those described in this paper, can certainly be useful, but to do this properly users must consent to having their interaction recorded and used to model their usage patterns. Our study has been useful for describing just how variable search interaction can be, identifying two classes of user, and suggesting three of the important dimensions in determining between-user variance (i.e., forward and backward motion, the branchiness, and time taken to traverse it). These dimensions can be useful to represent the interaction variance of the search population, and incorporated into a model for estimating the variability of a user's interaction with respect to that population.

Obtaining information that would help estimate query variability for at least the most popular queries should not be as difficult as estimating user variability. A measure describing how variable the interaction is for each query can be computed offline based on the interactions of a willing set of users and the two query variance dimensions we identified. This information could then be fed directly into the search engine at query-time, and used to select an appropriate form of interface support or ranking algorithms.

Assuming that we have information about the variability of interaction behavior, we can offer a range of search support options to support navigators and highly navigational queries, and explorers and highly exploratory queries. Personalizing search results [e.g., 30] is to be one way to address variability, but the following options are also available:

Navigational (Navigators or highly navigational queries)

Tools to support these users and queries will facilitate rapid access to information targets. Options include:

- *Teleportation:* Navigators and navigational queries were generally characterized by short, directed search trails. Teleportation [36] is a strategy that involves users jumping directly to their information target, with no steps in-between. Based on analysis of the intersection between multiple search trails frequently-visited destinations for a given query could be identified and offered to users as a list of search "shortcuts" to get them to their destination faster.
- *Personal Search Histories:* Previous searches (and perhaps search trails) can be stored for each user, and presented to them on the homepage of the search engine to support rapid navigation.
- *Interaction Hubs:* Navigators appeared to rely on important pages within domains to effectively perform aspects of their search. Surfacing these domains may give users branching points from which to pursue different aspects of their task.

Exploratory (Explorers or highly exploratory queries)

Tools to support these users and queries will facilitate browsing, understanding, and topic coverage:

- *Guided Tours and Domain Indices:* Explorers generally visited multiple domains when exploring. For the most popular queries, a list of "must see" domains could be constructed and presented to the user in some sensible order as a guided tour [38], or as a list of potentially relevant domains in a index accessible at all times during the search session.
- *Predictive Retrieval:* While not explicitly modifying the user interface, smarter caching and predictive retrieval using Web query logs or reconnaissance agents [19], could complement exploration activities. Predictive information can be used to open tabs for pertinent queries in addition to pre-fetching relevant links.
- *Support for Rapid Revisitation:* Our analysis showed that branchiness was an attribute of exploration. The history mechanism in the browser could be enhanced using recorded information such as query-terms, dwell times, and commonly selected branching information [34]. Enhanced back buttons can be added to the browser to return users to branch points or result pages [22].

More data than was available in this study is required to recommend the support to offer at the intersections between these two dimensions (e.g., explorers pursuing navigational queries). However, since it our intuition is that the nature of the query can outweigh the nature of the user (a proposition supported by [4]), it would be better in cases of doubt to tailor support to query variability rather than user variability. If there is no mechanism to automatically determine what support to offer, then systems should at least provide an interactive search toolkit, with clear descriptions of the circumstances under which each tool should be used.

In this section we have intentionally focused on extreme users and queries. The rationale has been to use features of their interaction to characterize exploratory and navigational behaviors, and in turn offer design recommendations. We are not proposing that the default search experience change for the average user (i.e., they will still be shown the same interface as at present), but rather that they be provided with the tools we recommend above, and perhaps assisted in selecting the appropriate tools by automated or human guides employed by the search engine. Extreme users may wish to set their defaults to these options, but it would not be required for them to do so. Tailoring support in this way has the potential to make Web search more inclusive. This has the benefit that all users can be empowered with new ways to search without having to use them, previously neglected extreme users have a way to meet their information-seeking objectives.

5. CONCLUSIONS

In this paper we have described a large-scale longitudinal log-based study investigating the levels of behavioral variability in users engaged in Web search activities. Over two thousand participants took part in the study over a five month period. The findings suggest that there are users and queries whose interaction is particularly consistent and particularly variable, and that it is possible to characterize many features of variation with a small number of underlying dimensions that could be useful in interface design. In addition to supporting the masses, who appear to interact fairly consistently, Web search engines should provide tools that are based on the needs of users that exhibit extreme search behaviors regularly. In future work we plan to create a catalog of Web-use patterns that goes beyond what we have proposed in this paper, and use those patterns to predict and perhaps explain the behavior of Web searchers.

6. REFERENCES

[1] Bates, M. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13: 407-424.

[2] Bederson, B.B. & Shneiderman, B. (2003). *The Craft of Information Visualization: Readings and Reflections*. Morgan Kaufmann.

[3] Bhavnani, S.K. (2001). Domain-specific search strategies for the effective retrieval of healthcare and shopping information. In *Proc. CHI 2002*, 610-611.

[4] Buckland, M.K. & Florian, D. (1991). Expertise, task complexity, and artificial intelligence: A conceptual framework. *J. Amer. Soc. Info. Sci*, 42 (9), 635-643.

[5] Card, S.K. *et al.* (2001). Information scent as a driver of web behavior graphs: Results of a protocol analysis method for web usability. In *Proc. CHI 2001*, 498-505.

[6] Catledge, L.D. & Pitkow, J.E. (1995). Characterizing browsing strategies in the world wide web. *Computer Networks and ISDN Systems*, 27(6): 1065-1073.

[7] Chi, E., Pirolli, P., Chen, J., Pitkow, J. (2001). Using information scent to model user information needs and actions on the web. In *Proc. CHI 2001*, pp. 490-497.

[8] Cutrell, E. *et al.* (2006). Fast, flexible filtering with Phlat - Personal search and organization made easy. In *Proc. CHI 2006*, 261-270.

[9] Dillon, A. & Watson, C. (1996). User analysis in HCI: the historical lesson from individual differences research. *International Journal of Human-Computer Studies*, 45(6): 619-637.

[10] Egan, D. (1988) Individual differences in human-computer interaction. In: *Handbook of Human-computer Interaction*. Elsevier, 543-568.

[11] Eisenstein, J. & Rich, R. (2002). Agents and GUIs from task models. In *Proc. IUI 2002*, 47-54.

[12] Ford, N. *et al*. (2002). Information seeking and mediated searching. Part 4. Cognitive styles in information seeking. *JASIST*, 53(9), 728-35.

[13] Hölscher, C. & Strube, G. (2000). Web search behavior of Internet experts and newbies. *Computer Networks*, 33, 337-46.

[14] Huberman, B. *et al.* (1998). Strong regularities in World Wide Web surfing. *Science*, 280 (5360): 95–97.

[15] Jansen, B.J., Spink, A. & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the Web. *Info. Proc. & Mgt.*, 36: 207-227.

[16] Jones, R. *et al.* (2006). Generating query substitutions. In *Proc. WWW 2006*, 387-396.

[17] Lau, T. & Horvitz, E. (1999). Patterns of search: Analyzing and modeling web query refinement. In *Proc. UM 1999*, 119-128.

[18] Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, 10(8):707–710

[19] Lieberman, H.L., Fry, C. & Weitzman, L. (2001). Exploring the Web with reconnaissance agents. *Communications of the ACM*, 44(8): 69-75.

[20] Malone, T.E. (1983). How do people organize their desks? *ACM TOIS*, 1(1): 99-112.

[21] Marchionini, G. (1995). *Information seeking in electronic environments.* Cambridge University Press.

[22] Milic-Frayling, N. (2004). SmartBack: Supporting users in back navigation. In *Proc. WWW 2004*, 63-71.

[23] Neilsen, J. (1993). *Usability Engineering*, Cambridge MA: Academic Press.

[24] Newell, A. & Simon, H. (1972). *Human Problem Solving*. Prentice-Hall.

[25] O'Day, V. & Jeffries, R. (1993). Orienteering in an information landscape: how information seekers get from here to there. In *Proc. CHI 1993*, 438-445.

[26] Pask, G. (1976). Conventional techniques in the study and practice of education. *British Journal of Educational Psychology*, 46, 12-25.

[27] Pirolli, P. & Card, S.K. (1995). Information foraging. *Psychological Review*, 106, 643-675.

[28] Pirolli, P. & Fu, W. (2003). Snif-act: A model of information foraging on the World Wide Web. In *Proc. UM 2003*, 45-54.

[29] Pitkow, J. & Pirolli, P. (1999). Mining longest repeating subsequences to predict World Wide Web surfing. In *Proc. USENIX Symposium*, 139-150.

[30] Pitkow, J. *et al.* (2002). Personalized search. *Communications of the ACM*, 45(9): 50-55.

[31] Rich, E. (1989). Stereotypes and user modeling. In *User Models in Dialog Systems*. Springer.

[32] Rose, D. E. & Levinson, D. (2004). Understanding user goals in Web search. In *Proc. WWW 2004*, 13-19.

[33] Russell, D.M. *et al.* (1993). The cost structure of sensemaking. In *Proc. CHI 1993*, 269-276

[34] Tauscher, L. & Greenberg, S. (1997). Revisitation patterns in World Wide Web navigation. In *Proc. CHI 1997*, 399-406.

[35] Teevan, J. *et al*. (2006). History repeats itself: Repeat queries in Yahoo's logs. In *Proc. SIGIR 2006*, 703-704.

[36] Teevan, J. *et al.* (2004). The perfect search engine is not enough: A study of orienteering behavior in directed search. In *Proc. CHI 2004*, 415-422.

[37] Teevan, J. *et al*. (2005). Beyond the commons: Investigating the value of personalizing web search. In *Proc. PIA 2005*.

[38] Trigg, R.H. (1988). Guided tours and tabletops: tools for communicating in a hypertext environment. *Transactions on Information Systems*, 6(4): 398-414.

[39] Weinreich, H., Obendorf, H., Herder, E. & Mayer, M. (2006). Off the beaten tracks: Exploring three aspects of web navigation. In *Proc. WWW 2006*, 133-142.

[40] Wexelblat, A. & Maes, P. (1999). Footprints: History-rich tools for information foraging. In *Proc. CHI 1999*, 270-277.