Understanding the Relationship between Searchers' Queries and Information Goals

Doug Downey
University of Washington
Seattle, WA 98195
ddowney@cs.washington.edu

Susan Dumais, Dan Liebling, Eric Horvitz
Microsoft Research
Redmond, WA 98052
{sdumais, danl, horvitz}@microsoft.com

ABSTRACT
We describe results from Web search log studies aimed at elucidating user behaviors associated with queries and destination URLs that appear with different frequencies. We note the diversity of information goals that searchers have and the differing ways that goals are specified. We examine rare and common information goals that are specified using rare or common queries. We identify several significant differences in user behavior depending on the rarity of the query and the destination URL. We find that searchers are more likely to be successful when the frequencies of the query and destination URL are similar. We also establish that the behavioral differences observed for queries and goals of varying rarity persist even after accounting for potential confounding variables, including query length, search engine ranking, session duration, and task difficulty. Finally, using an information-theoretic measure of search difficulty, we show that the benefits obtained by search and navigation actions depend on the frequency of the information goal.

Categories and Subject Descriptors
H.3.3 [**Information Search and Retrieval**]: Search process, Query formulation

General Terms
Experimentation, Human Factors. Measurement

Keywords
Web search, Information goal, User behavior

1. INTRODUCTION
When searching the Web, users typically issue a query to a search engine, are presented with a list of results, and then may click on one or more results in an attempt to satisfy an information goal. Along the way, the searcher may modify their initial query in various ways [14][16]. People have mixed experiences with this process: search engines do a great job of returning relevant results for some queries and a poor job for others. Although there has been a good deal of work characterizing the heavy-tailed Zipf distribution of both queries and target URLs [18][24], there has been much less work on understanding searchers' experiences and behaviors associated with these characteristics. Do search engines and searchers behave differently on rare queries than on common ones? Do rare queries represent rare informational goals, or simply atypical means of specifying common goals? How can answers to such questions guide research toward enhancing Web search experiences?

Search queries are the articulation of a person's information goals. People employ a mixture of search and navigation strategies to satisfy these goals. In laboratory studies, participants can be given known search goals or probed about their own information goals. In large-scale log studies, information goals must be inferred from patterns of user interactions. One approach to inferring searchers' information goals is to consider the patterns of pages viewed in sessions as well as the dwell times on pages as implicit indicators of interest [1][8][26][29]. For example, a researcher may start a session by searching for a another researcher by name, select a link to that colleague's home page from the results, navigate from the home page to a list of publications, and finally click a link to the paper which satisfies the intent of the search. In this paper, we shall take the last URL visited in a search session as a candidate proxy for a searcher's underlying information goal. Using such a proxy, for users' goals, we can identify relationships between queries and goals that are consistent across many users and diverse search tasks.

We seek to understand the relationship between the articulation of a goal (as represented by the search query) and the actual information goal (as represented by URLs visited in a session). We investigate how search behavior varies across rare and common queries, in the face of rare and common information goals. We present results from a large-scale log study of search sessions from users of major Web search engines.

After a discussion of previous work, we describe the data collection and session extraction methodology in detail. We then present experimental results describing user behaviors for sessions consisting of queries and target URLs of different frequencies. In some cases, we provide empirical confirmation of relationships one might expect to hold between goals and queries; in other cases, the findings are more surprising. Our key results include:

1) User behavior following a query varies significantly with the rarity of the user's query or destination URL.

CIKM'08, October 26-30, 2008, Napa, California, USA.

Web search engines are less effective as queries or target URLs become rarer; searchers are less likely to click results, and are more likely to reformulate queries (Sections 4.1 and 4.2).

2) A user's query is often much more specific or general than their underlying information goal. This has important implications for search engine effectiveness: search success is more likely when the relative frequency of the query *matches* that of the need (Section 4.3).

3) The decreased efficacy of search engines on rare needs can be characterized by increases in session length. The average number of queries the user must execute to satisfy rare goals is higher than for common goals (Section 4.4).

After establishing the results above, we proceed to examine potential confounding variables in the analysis. A number of different variables are known to correlate with user behavior as well as with the rarity of queries and information goals. We find that goal rarity appears to be the paramount influence in our findings, based on an investigation of a variety of variables including query length, search engine ranking, session length, and task difficulty.

Lastly, we discuss results showing that users often compensate successfully for decreased search engine effectiveness on rare queries by issuing more general queries, and then following hyperlinks to satisfy their need. We conclude with a discussion of the implications of our results for efforts toward enhancing Web search.

2. PREVIOUS WORK

Analyses of user behavior in a search engine context rely on a variety of methods including: user studies [2][11][21], offline query log analysis [13][16][24][25], and deeper instrumentation of session behavior using browser toolbar plug-ins [6][7][28][29]. User studies provide a detailed understanding of the behaviors and intentions of individuals. Toolbar data, on the other hand, provide a rich source of information about interaction patterns for a wide range of different users and tasks. Grimes et al. [10] describe the complementary relations among these different sources of information about user search behavior.

In this paper, we focus on large-scale analyses of logs data voluntarily shared by users via a browser toolbar, aimed at understanding both the diversity of information goals that searchers have and the queries they use to express those goals.

The information science literature provides several frameworks for understanding information seeking behaviors and processes [3][12][19]. People often begin an information-seeking episode with an "anomalous state of knowledge" (ASK) and seek information that will resolve the knowledge gap. The information seeking process involves a person translating their *information need* into a query, receiving a ranked list of results, and exploring the results via navigation among the results. If the initial search results do not satisfy the user's information need, the user may reformulate his or her query. Formulating and reformulating queries can be a challenging problem because the information goal may be unclear (*i.e.,* the user has uncertainty about the need), the user may introduce semantic or syntactic errors into the query, or the need may be unsatisfiable with the current resources [22]. Studies of how searchers modify their queries over the course of a session show that specialization of the query occurs more frequently than generalization [15], and that specialized queries tend to be longer [16]. Longer queries are more likely to be less frequent, so users often progress from frequent to rarer queries during a search session.

In the context of the Web, Broder [4] and Rose and Levinson [23] described three classes of user goals in search: informational, navigational, and resource or transactional. Lee et al. [17] used the distribution of clicks in search engine results and properties of the anchor text content to automatically identify navigational queries. Chang et al. [5] used analyses of clicked results' snippets to classify search queries as navigational or informational. Informational goals are similar to topical queries in the information retrieval literature. Navigational and transactional searches are more specific to the Web in which the user's goal is to find a Web site or to perform some Web-mediated activity.

Navigational searches are particularly interesting because they illustrate the distinction between the articulation of a user's goal (the query) and the user's underlying information goal. A query like *kelly blue book* is classified as a navigational query since it "demonstrates a desire by the user to be taken to the home page of the institution or organization in question" [4]. However, many searchers will navigate from this home page to other pages in this or other sites (*e.g.,* to look up the value of a particular used car; to post a car for sale; or to read reviews and ratings). These subsequent actions may be better characterizations of the searcher's information goal than their original query. Teevan et al. [26] and White and colleagues [28][29] have also noted that searches are just the starting point for richer information interactions that evolve over the course of a session. In the work reported here, we examine several ways of characterizing the information goal in a session.

Although there has been much written about the long tail of queries, less is known about the quality of search results and user interactions following the issuing of rare queries. Downey, Dumais, and Horvitz explored user behaviors associated with rare and common queries [6]. They found that users were more likely to click a result for common queries and conversely that they were more likely to reformulate rare queries. We extend this prior work on rare and common queries in several directions. We examine user behavior for both queries and target URLs that vary in frequency using much finer distinctions. We also examine how the correspondence between the expression of the goal (queries) and the target goal (last URL visited) influences user success.

3. METHODS

3.1 Data

We obtained fully-anonymized logs of URLs visited by users who opted in to provide data through a widely distributed browser toolbar. Each entry in the raw data consists of a unique machine identifier, a timestamp, the URL visited, and the URL's referrer (if any). Intranet and secure (https) URL visits are excluded at the source. In order to remove variability caused by geographic and linguistic variation in search behavior, we include only entries generated in the English speaking United States ISO locale. The results described in this paper are based on a sample of two weeks' worth of URL visits during June, 2006, representing more than 80 million URL visits from more than 206,000 unique users.

As is the case with many large-scale log analyses, if more than one person uses the same computer, they will have the same machine ID. However, we believe that it is reasonable to assume that a single session (defined below) represents the actions of a single user. Thus, we refer to a toolbar instance as a *user*.

3.2 Sessions, actions and goals

From the stream of URLs, we extracted queries issued to major Web search engines. We used HTTP referrer information to deduce whether a URL that followed a search was from a click on a results page or not.

3.2.1 Session definition

To create a user session, we ordered each user's activities by time. Each session begins with a search query and ends according to a subset of the criteria used by White and Drucker [28]. Specifically, we used 30 minutes of inactivity, a visit to the user's home page, or a login to a secured site (including mail, myspace and del.icio.us) to identify session boundaries. We used the page that appeared most commonly after a new browser window as a heuristic for identifying a person's homepage. We also experimented with defining session endpoints using just 30 minutes of inactivity. A manual inspection of several hundred queries found that the final URLs observed using the White and Drucker heuristics were generally more relevant to the search query that initiated the session than using a timeout heuristic alone. Because we were interested in understanding searcher's information goals, we only considered sessions that contained at least one click on a search engine result and that also ended with a visit to a Web page. Using this definition, we examined over 523,000 sessions from more than 105,000 unique users.

Table 1. Sample session

t_0	Query	peanut butter
t_1	SERP Click†	http://en.wikipedia.org/wiki/Peanut%20Butter
t_2	URL Visit	http://en.wikipedia.org/wiki/Sandwich
t_3	Requery	peanut butter sandwich recices
t_4	Requery†	peanut butter sandwich recipes
t_5	SERP Click†	http://pbsrecipes.com/lowfat.html
t_6	SERP Click	http://cooking.com/recipes/peanut_butter.html
t_7	URL Visit	http://cooking.com/recipes/pbsandwich.html

Actions marked with daggers (†) are post-query actions, As proxies for information goals, we examine the last SERP Click (t_6) and the last URL Visit (t_7).

Table 1 shows a sample session. The acronym SERP refers to the "search engine result page," thus a SERP Click is a click on one of the search engine results. The searcher begins at time t_0 by querying for *peanut butter*. They then click on a Wikipedia link presented on the results page. Once on Wikipedia, they navigate to a related topic. Shortly afterward they issue another query, but misspell the word *recipes*. Their next query corrects this typo, and they then click on two links on the result page. At t_7, the user navigates to another page on the same site as the previous result click. From this example, we may infer that the user was looking for peanut butter sandwich recipes.

3.2.2 Post-query actions

From the session data we explore the relationship between a query and the action immediately following the query. We use the relative frequency of these actions as a way to characterize the quality of the results for the query.

The actions of interest are:

- **SERP Click** – a click on the result page.
- **Requery** – a query reformulation.
- **URL Visit** – a visit to a Web page that is not a SERP Click or search engine query.

In Table 1, the post-query actions of interest are marked with daggers (†) – a SERP click at t_1, a Requery at t_4 and a SERP click at t_5. Figure 1 shows examples of the types of events considered in the processed data.

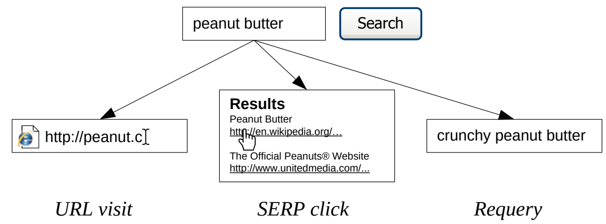

Figure 1. Post-query events

3.2.3 Information goals

We also explored how query behavior changes with respect to the rarity of information goal, as represented by the last URLs explored by the searcher. A user's search query is the *expression* of their information goal. However, this may not correspond to the user's actual information goal. Without conducting detailed interviews with searchers it is difficult to know precisely what they were searching for and whether they were satisfied with the results. We attempted to infer the nature of the user's information goal by examining the URLs that they visited near the end of the session. Others have looked at richer models of search interactions to evaluate the relevance of individual search results to the query [1][8]. Our interest is primarily in characterizing the information goals of searchers, and being able to do so on a very large scale to understand the wide variety of searcher's goals. We examine properties of URLs near the end of a session to do this. White et al. used a related notion of search destinations in their work [29] and showed how interactive search could be improved by suggesting common destinations for issued queries. We use similar ideas to characterize information needs.

We examined both the last SERP click in a session and the last URL in a session (which could be either a SERP click or a subsequent navigation) as proxies for the information goal. The *last URL visited* in a session has been used in prior research as a proxy for the user's intended destination [29]. Assuming that the last URL satisfies the goal of a given query is a simplification for several reasons: the goal may not have been satisfied; other pages in the session may have been necessary or sufficient to fulfill the goal; or (because detection of session boundaries is imperfect) the last URL may have been directed at a new goal. Nonetheless, the last URL visited is a reasonable proxy for the user's goal and, though approximate, such a definition is amenable to large-scale data analysis.

We additionally consider the *last SERP click* as a proxy. The last SERP click has the disadvantage of being restricted to the options that current search engines make available, and it ignores subsequent navigation. However, in manual inspection of user interactions we found that the last SERP click was sometimes more closely related to the query than the last URL, which is not surprising if search engines are doing a good job of ranking results. The SERP click also has the advantage of being available to search engines (*i.e.*, it does not require the client-side instrumentation we employ in this paper). Further, by comparing behavior *across* the two proxies, we can isolate the effects of post-SERP navigation (see Section 6).

One potential concern about using SERP clicks as a surrogate for information need is that users of tabbed browsers may load several SERP results into different tabs without actually reading them. We examined how often multiple URL loads occurred within one second of each other and found that this occurred in only 1.4% of the sessions. So we do not believe that this is a serious concern with the data used in our studies.

In summary, we examine two complementary methods for characterizing the user's information need, which we believe provide a rich picture of interactions with search engine results and subsequent navigation patterns. The two methods we use to characterize information goals are:

- **Last SERP Click–** the last SERP click in a session.
- **Last URL Visit –** the last URL visited in a session.

In the session illustrated in Table 1, the last SERP click in the session is shown at t_6, and the last URL visit in the session is shown at t_7.

3.3 Query and URL Frequencies

Search queries and URL accesses are known to follow a heavy-tailed Zipf distribution [24][25]. In the head of the distribution, a small number of queries and URLs are observed very frequently. However, there are many more queries and URLs in the tail that are observed much less frequently. We classified each query and URL in a session as *tail* or *non-tail,* and also considered finer-grained classifications of frequencies of occurrence.

We start with a methodology similar to that proposed by Downey et al. [6] to measure the frequency of occurrence. We use the first week of data to cumulate information about the frequency of queries and URLs, and the second week of data to measure user interaction patterns. This definition allows the results of these analyses to be applied in an online, sliding-window fashion. *Query frequencies* are obtained by counting the queries issued to major search engines. This is different than using corpus statistics to characterize the frequency of terms in queries. Since we are interested in searcher behavior and have access to query logs, the definition that focuses on searchers (and not authors) is preferable. *URL access frequencies* are obtained by counting Web page visits, regardless of whether these visits are in response to a query or not. We do so because we are interesting in understand common goals, and not just destinations that are easily accessible using search engines. To allow for comparison with earlier work by Downey et at. [6], we define the *tail* to contain queries or URLs observed in the second week of data that were not observed during the first week, and all other queries or URLs are *non-tail*. We also analyze finer-grained frequency bins using the frequency observed during the first week.

4. USER BEHAVIOR FOR RARE VERSUS COMMON QUERIES AND GOALS

We now focus on our primary experimental results showing differences in the behavior of searchers with changes in the rarity of queries or goals. We start by examining user behavior following individual queries, and then broaden the analysis to consider sessions of multiple queries.

4.1 Query frequency

We begin by examining whether searchers behave differently following rare and common queries. For each query in the session, we summarize the proportion of times that the next action is another query, a SERP click, or a URL visit. Since all of our sessions end in either a SERP click or a URL visit, we do not consider sessions that contain only unclicked queries.

Table 2. Post-query action by query frequency

	Post-query action		
Query frequency	**SERP click**	**URL visit**	**Requery**
Tail	0.579	0.064	0.357
Non-tail	0.725	0.069	0.207

Table 2 shows the distribution of post-query events broken down by the rarity of the query. Following tail queries, SERP clicks are less common (0.579 vs. 0.725) and requeries are more common (0.357 vs. 0.207), both of which indicate that the results returned by the search engine were not as useful to searchers. Z-tests for differences in proportions verify that these differences are statistically reliable. For these and all other proportions presented in the text, the results are significant at the .01 level, unless otherwise noted.

This trend also holds when we consider post-query actions within finer-grained frequency bins. Figure 2 shows the proportion of post-query actions as a function of finer-grained query frequency. As in the previous analysis, "tail" queries are defined as those that did not occur during the first week. The other frequencies are logarithmically binned. That is, queries that occur 1 to 9 times are in the first bin, those occurring 10-99 times in the next bin, etc. Figure 2 shows that SERP clicks increase and requeries decrease smoothly as the frequency of queries increases.

These results agree qualitatively with and extend the results reported by Downey et al. [6], and suggest that search engines are not doing as good a job of satisfying searchers for rarer queries as they do on more common ones.

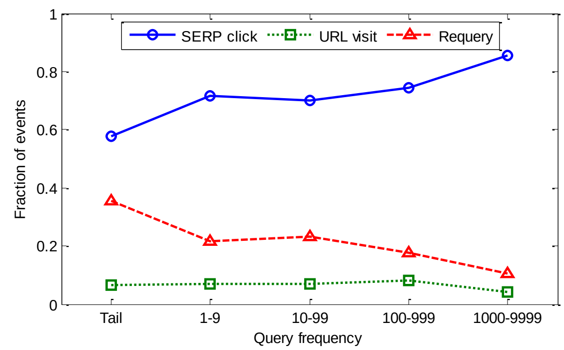

Figure 2. Post-query action by query frequency bins

4.2 Goal frequency

We now consider the influence of the rarity of the information goal. As described above, we examined two heuristics for characterizing searchers' information goals: one which uses the last SERP click in the session (Last SERP Click), and the other which uses the last URL in the session, which could be a SERP click or a subsequent navigation (Last URL Visit). Results for both are shown in the following tables.

Table 3 shows the effect of information goal rarity on post-query behavior. Using the last SERP click as the proxy for information goal, we observe behavior very similar to that shown in Table 2; SERP clicks are less likely for tail goals (0.590 vs. 0.667), and users are more likely to requery for tail goals (0.343 vs. 0.270). When we use the last URL visit as the proxy for informational goal, the pattern is similar but not as strong. Users are somewhat less likely to click (0.616 vs. 0.630) and more likely to requery for tail goals (0.324 vs. 0.296). There is also a small increase in URL visits following a query for non-tail URLs (0.074 vs. 0.060). These visits can come from a wide variety of sources (*e.g.*, typing directly into the address bar, navigating to bookmarks, etc.) and are difficult to interpret without more detailed analysis of the URLs.

Table 3. Post-query action by information goal frequency

		Post-query action		
Info.Goal	**Last URL frequency**	**SERP click**	**URL visit**	**Requery**
SERP Click	*Tail*	0.590	0.067	0.343
	Non-tail	0.667	0.063	0.270
URL Visit	*Tail*	0.616	0.060	0.324
	Non-tail	0.630	0.074	0.296

These trends are also evident when we consider post-query actions with finer-grained bins. Figures 3a and 3b show the proportion of post-query actions as a function of logarithmic frequency bins. Figure 3a shows the data when SERP Clicks are used to define user goals, and Figure 3b shows the same data for URL Visits. The reasons why post-query actions are less dependent on URL frequency in Figure 3b is not clear. In examining search sessions by hand, we have sometimes seen that the last URL visited in a session is unrelated to the query (particularly when the last URL is common), so this proxy for informational goals may be noisier. When using SERP clicks to characterize goals, however, we see very clearly that search engines perform less well at satisfying searchers for rarer information goals.

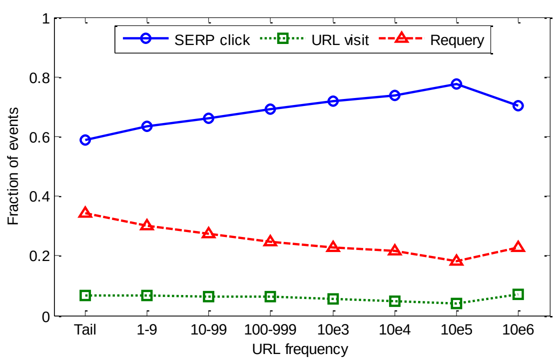

Figure 3a. Post-query action by last SERP click frequency bins

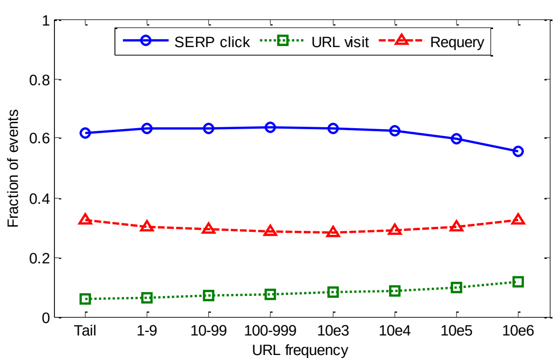

Figure 3b. Post-query action by last URL visit frequency bins

4.3 Frequency alignment

We have demonstrated that the distribution of post-query actions changes depending on the rarity of the query and information goal in isolation. We now examine the joint influences of rarity of query and information goal on search sessions. This combination allows us to examine the relationship between query articulation and the underlying information goal (as represented by destination URLs). Such analyses highlight the rich interactions and potential dependencies among the expression of goals and the goals themselves; searchers may formulate rare or common queries for rare goals—or common goals. By using the prior definitions of rare versus common queries and goals, we have four outcomes. Table 4 summarizes the results of studies of these four situations.

We observe that users are more likely to click a search result when the query and information goal are of similar frequencies. For tail queries, a SERP click is more likely when the information goal is also rare (0.583 vs. 0.570 for SERP Clicks and 0.589 vs. 0.560 for URL Visits). Likewise, results from common queries are clicked more for common information goals (0.771 vs. 0.627 for SERP Clicks and 0.741 vs. 0.706 for URL Visits). A similar pattern is observed for requeries, with queries being modified more when the query and URL frequencies do not agree. This suggests that search engines do a better job of identifying relevant results when the user's communication of their goal (the query) matches the frequency of their information goal (the URL).

Table 4. Post-query action by query and URL frequency

	Frequency		**Post-query action**		
Info. Goal	**Query**	**Last URL**	**SERP click**	**URL visit**	**Re-query**
SERP CLick	*Tail*	*Tail*	**0.583**	0.064	0.354
	Tail	*Non-tail*	0.570	0.065	**0.365**
	Non-tail	*Tail*	0.627	0.085	**0.289**
	Non-tail	*Non-tail*	**0.771**	0.061	0.168
URL Visit	*Tail*	*Tail*	**0.589**	0.058	0.353
	Tail	*Non-tail*	0.560	0.075	**0.365**
	Non-tail	*Tail*	0.706	0.066	**0.229**
	Non-tail	*Non-tail*	**0.741**	0.071	0.188

To further explore the alignment effect, we binned queries and URLs by their frequencies, again using logarithmic bins. Tables 5a and 5b show the SERP click probability for different frequency bins for queries (rows) and last URL (columns). The maximum of each column is shown in boldface type. We observe that the probability of a SERP click is maximized approximately across the diagonal. For each query frequency, SERP clicks are most frequent when the query and information goal match.We share several examples to help clarify the significance of this alignment. Consider a query like *webmd*, which is fairly common. Our log data shows that the most frequently clicked URL is also common: *http://www.webmd.com*. If searchers had the same information goal but described it in a less common way they would be less successful. Spelling errors (*e.g., webmb)* are an example of low frequency queries, and would require at least another click (on a query suggestion) or an explicit query reformulation to achieve success. Similarly, a query like *medical questions page* would not be the ideal way to get to the WebMD home page, although it may be all that a particular searcher can articulate about their goal. Low frequency information goals, on the other hand, tend to be satisfied best by low frequency queries. For example, if a searcher wants to get to an uncommon URL like *http://www.dc.state.fl.us,* the query *florida department of corrections* is a good way to get there. Conversely, a higher frequency query like *prisons* would be less successful and require more iteration to satisfy this information goal.

An important caveat to the alignment effect is that more common queries tend to be more effective than less common queries, regardless of goal rarity. This is shown to a small extent in Table 5a, and to a larger extent in Table 5b. In Table 5b the best performance for any row is obtained for common queries. We postulate that this effect results from users issuing more general queries than their specific need, and then "drilling-down" to satisfy their actual goal (commonly by using subsequent navigation beyond the SERP click, but occasionally by browsing deeper into search results). We investigate this phenomenon in more detail in Section 6.

To summarize, we have shown that searchers are more successful for common queries and common goals. We have also shown that regardless of query or goal frequency the best search success occurs when the relative frequency of the query *matches* that of the information goal.

Table 5a. SERP click probability by frequency bin using the last SERP Click as the information goal. (The column maximum is shown in bold.)

Last SERP Click Frequency	Query Frequency: *Tail*	*1-9*	*10-99*	*100…*	*1000…*
Tail	0.583	0.651	0.609	0.596	0.720
1-9	**0.597**	0.758	0.678	0.622	0.699
10-99	0.572	**0.788**	0.750	0.710	0.728
100…	0.523	0.772	**0.817**	0.776	0.796
1000…	0.476	0.650	0.814	0.878	0.823
10^4…	0.486	0.637	0.756	**0.906**	0.933
10^5…	0.477	0.548	0.588	0.716	**0.958**
10^6…	0.499	0.479	0.587	0.880	0.940

Table 5b. SERP click probability by frequency bin using the last URL Visit as the information goal. (The column maximum is shown in bold.)

Last URL Visit Frequency	Query Frequency: *Tail*	*1-9*	*10-99*	*100…*	*1000…*
Tail	**0.589**	0.705	0.680	0.722	0.839
1-9	0.585	**0.752**	0.700	0.725	0.812
10-99	0.561	0.744	0.729	0.743	0.821
100…	0.544	0.726	**0.751**	0.777	0.824
1000…	0.534	0.680	0.741	0.795	0.825
10^4…	0.535	0.679	0.698	**0.852**	0.893
10^5…	0.512	0.629	0.635	0.693	**0.940**
10^6…	0.510	0.623	0.630	0.654	0.842

4.4 Session characteristics

So far we have used the actions that follow queries to characterize search success, and shown that search engines perform more poorly on tail queries and URLs as measured by SERP clicks and requeries. We now examine in more detail how searchers move from initial queries to final URLs during the course of a search session.

Another measure of search engine performance is the number of queries a user issues during a session. We assume that the better a search engine is at satisfying the user's goal, the fewer queries will be required. For these analyses, we considered sessions that contained between 2 and 20 queries. Figure 4 shows the average number of queries in a session as a function of the frequency of the last SERP click. As the information goal becomes more common, the average number of queries in a session decreases from more than 4 queries to less than 2 queries.

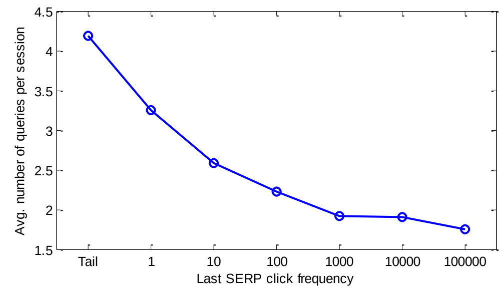

Figure 4. Average number of queries by frequency of information goal

Ideally, with each successive search in a session, users should come closer to fulfilling their information goal. We examine this by comparing the extent to which the frequencies of the first and last queries are aligned with the information goal. At the end of a search session, we expect that the frequency of the query should be closer to the frequency of the information goal than it was at the start of the session. Table 6 shows the proportion of sessions in which the frequency of the query (tail, non-tail) agrees with the frequency of the information need (tail, non-tail), for the first and last query in a session. The frequencies align 72.5% of the time for the last query compared with 66.5% for the first query. This trend holds for both tail and non-tail information goals. The lower overall level of agreement for tail URLs suggests that users may click on more popular results than their information need would suggest. However, it is important to note that the frequency of URLs in SERPs is influenced by the ranking provided by search engines, which may include a bias toward more common URLs.

The decreased search efficacy for rarer information goals is confirmed by the number of queries in a session. And, the alignment of query and URL frequencies is further supported by looking at changes in agreement during the course of a session.

5. ANALYSIS OF ALTERNATIVES TO QUERY AND GOAL RARITY

A number of different variables that are known to correlate with Web search behavior (e.g., the number of terms in a query, or the number of queries issued so far in a session) also correlate with query and goal rarity. An important question, therefore, is whether the behavioral variation presented above is captured equally well by these other variables, independent of query and goal rarity. In this section, we examine four such potential

Table 6. Proportion of sessions in which query and URL frequencies agree

Last SERP frequency	Agrees with: *First query*	*Last query*
All URLs	0.665	0.725
Tail URL	0.517	0.587
Non-tail URL	0.791	0.843

confounding variables, and find that rarity still remains a significant factor in explaining user behavior even after accounting for the other variables.

5.1 Role of query length

Phan et al. [21] have shown that query length is associated with the specificity of a user's information goal and found that longer queries were generally associated with more specific information goals. Prior work has also shown that longer queries tend to occur less frequently [13]. Thus, the differences that we have observed in the distribution of actions following a query may reflect how search engines respond to queries of different lengths rather than to queries of different frequencies. In this section, we present experiments that show this conjecture to be false – query rarity in fact has a more significant effect on behavior than does query length.

We examined post-query actions as a function of number of words in each query. Figure 5 shows the proportion of post-query actions that are requeries as a function of query frequency (top panel) and query length (bottom panel). The top graph shows that the probability of a requery drops (from almost 50% to less than 20%) as query frequency increases from the tail to 100+ occurrences. But this effect is very similar for queries of different lengths. The bottom graph shows that the probability of a requery is fairly flat as the number of words in the query increases, but varies substantially as query frequency changes. Similar results are observed for SERP clicks as well, but are not shown. Thus, the differences in distribution of actions following a query that we observed appear to be related to query frequency rather than query length.

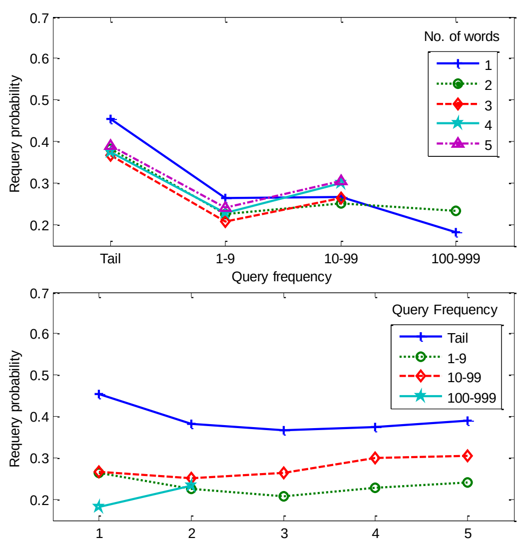

Figure 5. Proportion of post-query query actions that are requeries, by query frequency and query length

It is worth noting that much research in information retrieval shows that query length is positively correlated with search engine effectiveness for TREC-style ad hoc queries and relevance feedback experiments [27] and for interactive user studies [2]. Most experimental IR systems use probabilistic best-match techniques in which combining more terms in a weighted fashion leads to improved retrieval. In contrast, Web search engines use exact-match techniques in which all query words must match a Web page (in combination with a large number of other factors) for ranking. In Web search, we find that longer queries are not as successful as shorter queries. We believe that the differences between phenomena observed in traditional information retrieval and Web search arise because of the nature of the underling matching techniques. In addition, we found that the query frequency is more important than query length, indicating perhaps that Web search engines are optimized to handle common requests.

5.2 Role of search engine ranking

Tables 5a and 5b show that queries are more successful when their frequency matches that of the underlying goal. One candidate explanation for this finding is that search engines return results with frequencies matching that of the query. If so, then searchers will have few options but to choose SERP URLs that match the query in frequency. Here, we investigate whether the alignment observed in Tables 5a and 5b is simply due to the mix of result frequencies that search engines present.

Table 7 lists the fraction of presented SERP URLs that were clicked, for varying query and SERP URL frequencies. That is, for each cell, we report the proportion of presented URLs that were clicked. If the alignment in Tables 5a and 5b were due only to the mix of result frequencies presented by search engines, we would expect all values for each row in Table 7 to be the same independent of query frequency. On the contrary, the table illustrates that a result is more likely to be clicked, per presentation, when its frequency matches the query frequency. Thus, we have shown that the alignment in Tables 5a and 5b is not simply an artifact of search engine ranking, as it persists after accounting for how frequently results are presented.

Table 7. Ratio of the number of clicked result URLs to the number of presented result URLs at varying query and URL frequencies. (The row maximum is shown in bold.)

		Query Frequency				
		Tail	*1-9*	*10-99*	*100…*	*1000…*
URL Frequency	*Tail*	**0.052**	0.037	0.032	0.028	0.030
	1-9	**0.090**	0.067	0.047	0.028	0.015
	10-99	**0.128**	0.117	0.073	0.050	0.031
	100…	0.161	**0.206**	0.138	0.065	0.061
	1000…	0.182	0.242	**0.277**	0.179	0.075
	10^4…	0.204	0.230	0.278	**0.324**	0.205
	10^5…	0.161	0.211	0.216	0.360	**0.380**
	10^6…	0.129	0.239	0.331	0.307	**0.470**

5.3 Role of session length

In Section 4.4, we showed that as information goals become rarer, search sessions become longer. Previous results have shown that as additional queries are issued in a session, the likelihood of success on a query decreases [7]. Thus, one candidate explanation for the relatively lower success rate on queries for tail goals is simply that these queries tend to occur later in a search session, where success is less likely. In this section, we present experiments showing that this explanation is incorrect – in fact, the substantial differences in user behavior due to goal rarity occur *only* at the beginning of sessions.

We measured how a searcher's probability of success varies with the number of queries issued previously in a session. For this analysis we define a search to be successful if the user clicks on a SERP result and does not issue another query in the session. Figure 6 also shows the probability of success as a function of number of queries executed so far in the session. As additional queries are issued, each subsequent query has a lower probability of success. This result is intuitive: if an information goal has already required several queries, that goal is challenging and thus unlikely to be satisfied by a single additional query.

The figure shows that success on the *initial* queries depends heavily on the frequency of the last SERP URL. On the first query, searchers are 2.5 times more likely to succeed for common URLs (0.64 vs. 0.27). However, this difference disappears as session length increases, suggesting that long sessions tend to involve difficult information goals, regardless of the frequency of the last SERP URL. Thus, we have shown that the difference in user behavior observed for queries aimed at goals of varying rarity (**Error! Reference source not found.**) is not simply a consequence of the point in the session at which those queries were executed.

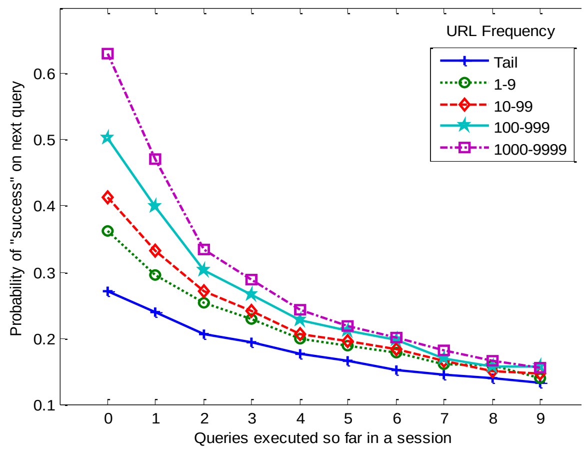

Figure 6. Success in terms of number of queries executed in a session

5.4 Role of search task difficulty

In Section 4.4 we showed that the number of queries required to achieve a user's information goal increases as the goal becomes rarer. But is this increase in user effort simply due to the increased difficulty involved in finding a more obscure target page, or is it instead due to ineffectiveness of search engines in handling rare queries?

We address this question with an information-theoretic analysis aimed at characterizing the difficulty of achieving the information goal [20]. Specifically, we define the difficulty of the search task as the expected number of bits of information required to identify the URL satisfying the goal, assuming an optimal code. Thus we characterize the difficulty of a search task as follows:

$$TaskDifficulty = \log_2 \frac{1}{P(Goal\ URL)}$$

where $P(Goal\ URL)$ is the frequency of the URL satisfying the user's goal, expressed as a probability. Using this definition, difficult tasks are those whose goals are rare. The Goal URL can either be a SERP Click or URL Visit, as described in Section 3.2.3. In these experiments, to produce probabilities we smooth the frequencies of the goal URLs using Good-Turing estimation [9]. We also define the *average information gain of a query* as the task difficulty divided by the number of queries in the session, and we define the *average information gain of a URL visit* similarly.

In Figure 7 we show how the average information gain of a query action varies with the rarity of the user's information goal. The curves show how information gain per query varies with goal rarity for our two methods of characterizing information goals. Using the last SERP Click as the user's goal, we find that queries are most effective for finding SERP URLs that are neither too common nor too rare, with a peak at a frequency between 100 and 999 visits in our data set (corresponding to a probability between 1.2×10^{-6} and 1.2×10^{-5}). The average information gain at the peak is about 50% higher than at either the rare or common extreme. The downward trend in information gain (for both curves) as needs become rare suggests that search engines are, in fact, less effective at returning URLs that satisfy rare goals, even after accounting for the difficulty of the goal. We describe the differences between the curves in more detail in the next section.

6. IMPACT OF GOAL DIFFICULTY ON SEARCH STRATEGY

One strategy users employ in Web search is to execute a more general query than their need, and then after clicking a SERP result navigate via hyperlinks to find a page that satisfies their goal [26][29]. In this section, we show that the results presented in Section 4.3 which suggested that this strategy is a particularly effective approach for rare goals, are confirmed using the information-theoretic analysis of search strategies described in the previous section.

Tables 5a and Table 5b summarize the frequency alignment result in which the best performance (for each column) is observed when the frequencies of the query and goal URL are similar. There are, however, some interesting differences between the tables. In Table 5b, the success rate is highest for more common queries independent of the goal rarity. That is, for all rows, the best performance is observed for the most frequent queries (right column). In Table 5a, this pattern is less consistent. Thus, when achieving the search goal using just SERP URLs (Table 5a), common queries are not uniformly the most effective. However, when the user can use navigation *beyond* the SERP to achieve this search goal (Table 5b), more common queries are more effective than rare queries independent of the goal rarity. In fact, using a more common query improves effectiveness by an average of 9% if subsequent navigation is employed (Table 5b), versus only 2% if not (Table 5a). This suggests that users can achieve greater success in search if they utilize more general queries than their

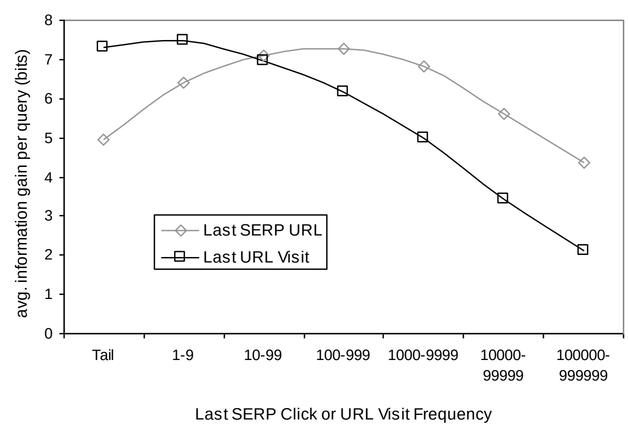

Figure 7. Average information gain of query actions as goal rarity varies

goal, and navigate to their goal page.

We confirm these findings using the information-theoretic analysis described in the previous section. Figure 7 shows how information gain per query varies when we use the last URL Visit and the last SERP URL to represent the goal of the session. The *differences* between the two curves show that additional navigation actions that occur in the URL Visit curve can increase the information gain per query for rare goals (those with frequencies less than 10), while decreasing the gain for common goals. Thus, the strategy of clicking a relatively general SERP URL and then navigating to a more specific goal page allows a user to increase the information they gain from each query in a session.

Of course, for this strategy to be effective, the number of navigations needed to obtain the goal must not be too large. How many URL Visits does the user have to execute in order to get the benefits for rare goals shown in Figure 7? On average, we found that a query provided about six times as much information gain as a URL Visit. However, queries and URL Visits have distinct patterns of effectiveness for goals of different frequencies. The average number of queries executed in a session increases substantially as goals become rarer, whereas the number of URL Visits stays relatively constant. The relative effectiveness of the two actions also varies with the length of the search session, as shown in Figure 8. In the figure, the information gain due to a single query (1 Query) is compared with ten times the information gain of a navigation action (10 URL Visits). Queries are especially informative for one-query sessions (that is, cases in which only a single query is required to satisfy the information goal, such as navigational queries), but queries decrease in effectiveness more rapidly than URL Visits as session length increases.

The information-theoretic measure of search difficulty quantifies the challenge inherent in search tasks and makes precise the rate at which queries and URL visits yield information regarding the target goal. Queries are most informative for goals which are neither too rare nor too common. For rare goals, the strategy of starting with a general query and navigating to more specific URLs is the most effective.

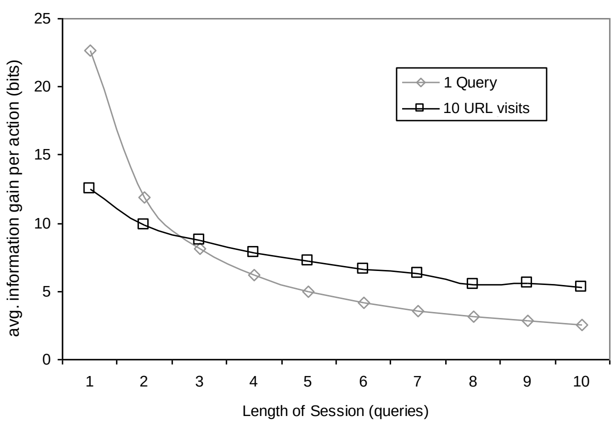

Figure 8. Information gained versus length of session

7. CONCLUSION AND FUTURE WORK

It is well known that search queries and URL accesses follow a heavy-tailed distribution, but less has been known about how the wildly varying commonality of queries and information goals influences user behavior or the performance of search engines. We presented the first large-scale log analyses that characterize the distribution of rare queries and goals in Web search. We also examined the influence of the rarity versus commonality of queries and goals on user interactions with today's search engines. Our analyses uncovered significant differences in user behavior for rare and common queries and URLs. Several measures of user interaction (results clicks, query reformulations and session length) indicate that search engines are doing a better job at satisfying common information goals. We also established that important differences exist between the *expression* of an information goal (as measured by the query) and the underlying goal itself (as measured by the last URLs visited in a search session). The best search success occurs when there is an alignment between the frequency of goals and expression of those goals.

We also validated our results by probing a set of potential confounding variables--including query length, search engine ranking, session duration, and task difficulty—and found that the differences in user behavior due to the rarity of queries and goals persist after accounting for these other variables. Lastly, using an information-theoretic analysis of user actions during a search sessions, we showed that clicking a general result and then navigating to more specific destination URLs can be an effective strategy for satisfying rare information goals.

Incorporating methods and machinery for addressing the long tail of rare queries promises to enhance Web search. Our findings on the link between the commonality of queries and goals and interactions with search engines are descriptive of searcher's behaviors with existing Web search engines. Nevertheless, we believe that the observations have implications for how to better support searchers using existing systems, and how to design new systems that will more effectively satisfy user's information needs. The alignment results, suggest that search engines might perform better if rankings took into consideration the query frequency. For example, results with similar frequencies to the query might be boosted in the ranking (i.e., common URLs could be ranked higher for common queries, and rarer URLs ranked higher for rare queries). This could be especially important for rare information goals, where search success is lower. Identifying rare goals in real-time is challenging technically, but could be accomplished by analyzing the patterns of queries, query reformulations and URL visits during the course of a search session (*e.g.,* by generalizing some of the techniques developed in [7]). Another direction for improvement would be to better support searchers in articulating their information goals. One way to do so would be to offer query suggestions that are guided by the frequencies of queries in the search session. Any such modifications to ranking or query suggestion algorithms would benefit from a better understanding of how query and URL frequencies change over time in response to news events, Internet memes, and so on.

There are also a number of interesting empirical and theoretical extensions of this work which we would like to examine. An important direction is to develop additional methods for automatically identifying searcher's information goals, and to improve the identification of sessions that are associated with the same information goal. We have explored using the final SERP Click or URL Visit to represent the information goal, but we could extend this by considering characteristics of domains in addition to URLs, or classifying URLs into a taxonomy like the Open Directory to provide a higher-level topical representation of search activities. We would also like to complement our large-scale log analyses with field studies to obtain a richer understanding of searcher's information goals and to develop predictive models along the lines of those proposed by Fox et al. [8] to link session outcomes and implicit behavioral patterns.

Another future direction involves understanding the dynamics of search sessions, including how searchers move from common to rare information goals or vice versa, in both their query reformulations and in the pages they visit. Finally, as search engines begin to incorporate user histories in rankings, there are a variety of interesting challenges in how best to combine individual and aggregate query and interaction history to support individuals in satisfying their rare or common information needs.

8. REFERENCES

[1] Agchtein, E., Brill, E., Dumais, S. and Ragno, R. Learning user interaction models for predicting web search results preferences. *Proc. SIGIR 2006*, 3-10.

[2] Belkin, N. J., Cool, C., Kelly, D., Lee, H.-J., Muresan, G., Tang, M.-C. and Yuan, X.-J. Query length in interactive information retrieval. *Proc. SIGIR 2003*, 205-212.

[3] Belkin, N. J., Oddy, R. N. and Brooks, H. M. ASK for information retrieval: Part 1. *Journal of Documentation, 38 (2),* 61-71, 1982.

[4] Broder, A. A taxonomy of web search. *SIGIR Forum, 36 (2),* 3-10, 2002.

[5] Chang, Y.-S., He, K.-Y., Yu, S. and Lu, W.-H. Identifying user goals from Web search results. *Proc. WWW 2006,* 1038-1041.

[6] Downey, D., Dumais, S. and Horvitz, E. Heads and tails: Studies of web search with common and rare queries. *Proc. SIGIR 2007*, 847-848.

[7] Downey, D., Dumais, S. and Horvitz, E. Models of searching and browsing: Languages, studies, and applications. *Proc. IJCAI 2007*, 2740-2747.

[8] Fox, S., Karnawat, K., Mydland, M., Dumais, S. and White, T. Evaluating implicit measures to improve the search experience. *Transactions on Info. Systems, 23(2),* 147-168, 2005.

[9] Gale, W. and Sampson, G. Good-Turing frequency estimation without tears. *Journal of Quantitative Linguistics, 2*, 217-237, 1995.

[10] Grimes, C., Tang, D. and Russell, D. Query logs are not enough. *Proc. WWW 2007 Workshop on Query Log Analysis.*

[11] Holscher, C. and Strube, G. Web search behavior of internet experts and novice. *Proc. WWW 2002*, 337-346.

[12] Ingwersen, P. and Järvelin, K. *The turn: Integration of information seeking and retrieval in context*. New York: Springer-Verlag, 2005.

[13] Jansen, B., Spink, A. and Saracevic, T. Real life, real users, and real needs: a study and analysis of user queries on the web. *Info. Proc. and Management, 36,* 207-227, 2000.

[14] Jones, R.,Rey, B, Madani, O. and Greiner, W. Generating query substitutions. *Proc. WWW 2006*, 387-396.

[15] Jones, R. and Fain, D. Query word deletion prediction. *Proc. WWW 2003*, 435-436.

[16] Lau, T. and Horvitz, E. Patterns of search: Analyzing and modeling Web query refinement. *Proc. User Modeling 1999*, 119-128.

[17] Lee, U., Liu, Z. and Cho, J. Automatic identification of user goals in Web search. *Proc. WWW 2005,* 391-400.

[18] Lempel, R. and Moran, S. Predictive caching and prefetching of query results in search engines. *Proc. WWW 2003*, 19-28.

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

[20] Mei, Q., Church, K. Entropy of search logs: How hard is search? With personalization? With backoff? *Proc. WSDM 2008,* 45-54.

[21] Phan, N., Bailey, P. and Wilkinson, R. Understanding the relationship of information need specificity to search query length. *Proc. SIGIR 2007*, 709-710.

[22] Pharo, N. *The SST Method Schema: A Tool for Analysing Work Task-Based Web Information Search Processes.* Ph.D. dissertation, University of Tampere (Finland), 2002.

[23] Rose, D. and Levinson, D. Understanding user goals in web search. *Proc. WWW 2004*, 13-19.

[24] Silverstein, C., Henzinger, M., Marais, H. and Moricz, M. Analysis of a very large AltaVista query log. *Technical Report 1998-014*. Digital SRC, 1998.

[25] Spink, A. and Jansen, B. *Web search: Public searching of the Web.* Kluwer Academic Publishers, 2004.

[26] Teevan, J., Alvarado, C., Ackerman, M.S. and Karger, D.R. The perfect search engine is not enough: A study in orienteering behavior in directed search. *Proc. CHI 2004*, 415-422.

[27] Voorhees, E. and Harman, D. (Eds.) *TREC: Experiment and evaluation in information retrieval.* MIT Press, 2005.

[28] White, R. and Drucker, S. Investigating behavioral variability in Web search. *Proc. WWW 2007*, 21-30.

[29] White, R., Bilenko, M. and Cucerzan, S. Studying the use of popular destinations to enhance web search interaction. *Proc. SIGIR 2007*, 159-166.