Understanding and Predicting Personal Navigation

Jaime Teevan, Daniel J. Liebling, and Gayathri Ravichandran Geetha

Microsoft Corporation
Redmond, WA, USA

{teevan, danl, garavich}@microsoft.com

ABSTRACT

This paper presents an algorithm that predicts with very high accuracy which Web search result a user will click for one sixth of all Web queries. Prediction is done via a straightforward form of personalization that takes advantage of the fact that people often use search engines to re-find previously viewed resources. In our approach, an individual's past navigational behavior is identified via query log analysis and used to forecast identical future navigational behavior by the same individual. We compare the potential value of personal navigation with general navigation identified using aggregate user behavior. Although consistent navigational behavior across users can be useful for identifying a subset of navigational queries, different people often use the same queries to navigate to different resources. This is true even for queries comprised of unambiguous company names or URLs and typically thought of as navigational. We build an understanding of what personal navigation looks like, and identify ways to improve its coverage and accuracy by taking advantage of people's consistency over time and across groups of individuals.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *query formulation, search process*.

General Terms

Human Factors, Measurement.

Keywords

Navigation, query intent, personalized search, Web search, personal navigation, re-finding, query log analysis.

1. INTRODUCTION

One common way that Web search engines are used is to navigate to particular information resources. For example, a person looking to buy a book on Web search and data mining may, instead of searching directly for a book on the topic, issue the query *amazon* in order to navigate to the Amazon.com website where a relevant book can then be identified and purchased. Over 25% of all queries are navigational in nature, according to an in situ survey of people actively searching the Web [5].

If search engines are able to identify that a query is navigational, and to identify the query's intended navigational target, they can use that information provide significant benefit to their users. At a most basic level, they can display the target in a prominent manner that is easy for users to find and select. Additionally, this can be done quickly via better caching for navigational queries [17], and the interface can be designed to help support the desired

intent by providing, for example, links directly into the site's content [8] or access to appropriate meta-data or site functionality. Search engines may also be able to provide their users with more appropriate advertisements [2].

Several approaches have been explored to identify navigational queries, including analysis of the query string (e.g., is the query a URL or company name [3, 4, 12, 20]?) and behavioral data (e.g., does everyone click on the same result after issuing the query [3, 11, 15, 16]?). But while some queries are used to navigate to a particular resource by all who issue them, there are many more queries with navigational intent where the intent or intended resource is not obvious, even when it seems like it should be from the query string. For example, the reader of this paper may use a search engine to navigate to the WSDM 2011 homepage via the query *wsdm*, while a person interested in country music in the Midwest may use the same query to navigate to the WSDM-FM radio station homepage. Others may not use the query *wsdm* for navigation at all, but rather issue it with an informational intent to learn more about Web Services Distributed Management.

To truly understand whether a particular instance of query is navigational requires understanding the individual user's intent when they issue it. We find it is possible to easily and accurately identify a significant portion of queries with navigational intent and the associated target by using an individual's past search behavior via an approach that we call "personal navigation." We identify personal navigational behavior once a user has used a query to navigate to a particular result twice before. For example, someone who has searched for *wsdm* several times and clicked on http://wsdm2011.org every time they did can be expected to click on the same result the next time they issue the query.

Personal navigation presents a real opportunity for search engines to take a first step into safe, low-risk Web search personalization. Most personalization approaches rely on explicit or inferred user profiles to guess what new content might be of interest to a user for a given query. Here we look at how to capture the low-hanging fruit of personalizing results for repeat queries. Our ability to reliably identify navigational intent for queries that appear informational suggests navigational behavior may be more common than previously believed. What is more, there is the potential to significantly benefit users with the identification of these queries, as the identified targets are more likely to be ranked low in the result list than typical clicked search results.

After a brief description of the query logs used for our analysis, we explore general navigational behavior where everyone is assumed to use the same query to navigate to the same result. We expose several flaws in this approach, and introduce personal navigation as an alternative. We present a straightforward algorithm for identifying personal navigation behavior, and show that many queries can be easily and accurately identified in this way. We explore how our ability to predict personal navigation is impacted by the consistency of the behavior over time and across individuals, and conclude with a discussion of how repeat behavior can be used to improve the search experience.

2. QUERY LOGS

To explore navigational behavior, we analyzed the query logs from the Bing search engine. From the logs, we sampled information related to approximately 70 million queries gathered from over 21 million users. For each query, the sample contained information about when the query was issued, who issued it, and the URL and rank of any clicked results. The sample was filtered to remove bots and spam, and processed so that pagination and back button clicks were treated as the same query. Only queries issued through to Bing via the Web interface were included.

Users were associated with an anonymous ID stored in a browser cookie during their first search. As is the case with most log analyses, if a person has more than one computer, that person will have multiple IDs. Conversely, if more than one person uses the same account on a computer, they are amalgamated into a single user. These IDs have varying life spans commensurate with the nature of Web browser cookies, and become less useful over longer periods of time. We used logs from users in the United States English language locale. The data included users from outside the U.S. who selected this locale preference.

Query strings were normalized by removing excess whitespace, converting the text to lowercase, and removing punctuation while preserving the *n*-grams for terms typically joined by a punctuation character (e.g., *facebook.com*). When discussing clicked URLs, we consider only those presented in the algorithmic section of the results page, and ignore clicks on advertisements, query suggestions, or other things. For rank analysis, links nested under a parent link (i.e., "deep links") in the algorithmic result section are considered to have the rank of their parent link.

3. GENERAL NAVIGATION

We begin our analysis of the query logs by looking at navigation behavior across all users. Following a discussion of related work, we present how we identified general navigational queries, look at what can be learned from these queries, and show that a single approach for all people can fail to capture navigational intent.

3.1 Related Work on General Navigation

Broder [5] developed a taxonomy of Web search queries based on the reasons why search engine users reported having issued a query. He identified three different search intents: navigational, informational, and transactional. *Navigational* searches were defined as those intended to find a particular Web resource. (*Information* searches are intended to find information on a topic, and *transactional* searches are intended to perform an activity.) One quarter (25%) of all queries have a navigational intent, according to an interstitial survey he conducted of Web searchers.

Despite the prominence of navigational queries, identifying them has proved challenging. One common approach has been to use the query string to identify queries with which there are particular resources clearly associated, such as queries consisting of company names (e.g., *amazon*) or URL fragments (e.g., *msn.com*). Using manual classification of 400 queries taken from a query log, Broder [5] identified 20% of all queries as navigational. Rose and Levison [20] narrowly defined navigational queries as ones where a user wants to be taken to the homepage of a specific institution or organization (including queries for companies and universities, but not, for example, for celebrity names). They manually classified 1500 queries and identified 11.7% to 14.7% of queries as navigational according to that definition. Jansen et al. [12] extended this work to automatically identify navigational queries based on features derived primarily from their query

strings, and found that 10% of queries were navigational. Kang and Kim [14] used information about the occurrence of query terms in Web documents (e.g., anchor text and part-of speech information) to automatically classify queries as navigational.

However, as we will show, many queries that appear from their text to unambiguously refer to a single resource are actually used in practice to find multiple different resources. Another approach to identifying navigational queries (and the one we employ in this section) is to use aggregate log data to identify consistent post-query click behavior. Researchers have used machine learning and features of the query string, results, and behavior to classify query intent along a variety of different schemes [3, 4, 11]. For example, Lee et al. [15] used click behavior to identify queries for which one result is particularly likely to be clicked. Lu et al. [16] further explored behavioral identification of navigational queries using machine learning and feature selection. They found that user click distribution features are the most important for identifying navigational queries. The success of these approaches has been reasonable, ranging from 50% [4, 15] to 80% [3, 11] accuracy, depending on the exact experimental setup.

In this section, we demonstrate that properties of the query string and aggregate click behavior are not sufficient to identify when an individual intends a query to be navigational. The aggregate behavior-based approach we use to identifying navigational behavior is similar to the ones described above, and allows us to accurately identify queries that represent a sizeable portion of search engine navigational traffic. But when we look closely at the aggregate queries identified, we find significant opportunity for improvement. In the subsequent section we show how we capitalize on this opportunity by using individual user data.

3.2 Identifying General Navigation

We automatically identified a set of general navigational queries from the query logs by looking for queries that were followed by everyone clicking the same result, as measured by the query's click entropy. Click entropy for a query, $CE(q)$, is calculated as:

$$CE(q) = -\sum_{u \in P_c(q)} p_c(u|q) * \log p_c(u|q),$$

where $P_c(q)$ is the collection of URLs clicked on for query q and $p_c(u|q)$ is the percentage of clicks on URL u among all clicks for query q. While high click entropy can be the result of many different factors (including how much the results presented for the query change and how many time people usually click following the query), low click entropy it is a good approximation of similar intents [28]. We identified queries with a click entropy lower than 1.00 during the first week in June 2010 as navigational.

Uncommon queries can have very low click entropy merely because the query has not been issued very often. For example, the query *health coverage* has a click entropy of 0.87, despite appearing informational. The low observed click entropy is probably a result of the fact that only 49 distinct users issue the query in our sample. To avoid misidentifying queries as navigational due to lack of data, we only considered queries to be navigational if they were very common, having been issued by more than 10,000 distinct users. But even popular queries can have low click entropy as a result of there being very few observed clicks for the query. Popular queries that are not followed by clicks are often ones where the searcher's query intent is met directly on the search result page. For example, 10,371 people in our sample issued the query *definition of*

Table 1. Query features, broken down by whether the query was used for general navigation or personal navigation, as compared with all search engine queries.

| | | All queries | Navigational | |
			General	Personal
String	Length (chars)	16.76	9.00	11.73
	Length (words)	2.70	1.46	1.84
	URL fragment	1.61%	18.96%	1.34%
Freq	Average frequency	659,773	3,295,214	245,992
	Issuances per user	2.04	2.15	4.59
Clicks	Clicks per query	0.82	0.83	1.11
	Click entropy	1.10	0.66	0.37
	Click rank	1.17	1.03	1.31
	Clicks on rank 1	91.38%	98.16%	84.55%

pococurante, and the query had a click entropy of 0.96. This relatively low click entropy is not a result of everyone navigating to the same resource, but due to the fact that a result was clicked following the query only 90 times because the definition is provided on the search result page. For this reason, we also only considered queries to be navigational if they produced at least a total of 1000 search result clicks.

3.3 Understanding General Navigation

In this way we identified 390 unique general navigation queries. A summary of some of the basic characteristics of these queries, as compared with the average search engine query, can be found in Table 1. (Features of personal navigation queries are also shown, and will be discussed later.) Averages are computed over all query instances, as opposed to over unique queries. Thus a query that is issued many times contributes more, for example, to the average query length than a query that is issued fewer times.

The most common general navigational queries were *facebook*, *youtube*, and *myspace*. Consistent with related work [3, 4, 12, 20], most of the general navigational queries were company or organization names, and many contained URL fragments. Only 1.61% of all search engine queries contained a URL fragment, but 18.96% of the general navigational queries did. Additionally, the general navigational queries were short, averaging only 1.46 words and 9.00 characters in length, which is considerably less than typical for Web search queries in general. This may reflect that fact that navigational queries are most useful when they are quick and easy for the user to call to mind and type.

As the general navigational queries were selected to be popular, it is not surprising that, although few in number, they accounted for 12.02% of the total query volume. On average, each unique general navigational query was issued over 3 million times, which is almost five times as much as a typical query. They also had, by definition, a more consistent click pattern than other queries did. Although a similar number of results were clicked following general navigational queries (0.83) as for Web search as a whole (0.82), the average click entropy was much lower. It averaged 0.66 for the navigational queries and 1.10 for all queries.

Our approach to identifying general navigation permitted us to make a fairly accurate prediction as to what would be clicked following an identified query. The result most commonly associated with a general navigational query was clicked following 71.74% of the queries for which there was a click. The

remaining 28.3% of the time, the query was used in some other way, such as to visit another result or to visit several results. As a comparison, the result most commonly associated with the average search engine query was clicked only 37.75% of the time. Another difference between the general navigational queries and all queries is that clicked navigational results tended to be ranked much higher. People clicked on the first result returned for a general navigational query for 98.16% of the time, compared with 91.38% of the time for Web search engine queries in general. The average click rank was 1.03, higher than the non-navigational click rank of 1.17. Overall, the search engine did a particularly good job meeting users' needs for general navigational queries.

Nonetheless, identifying navigational queries via aggregate query behavior is imperfect. The strong restrictions we imposed in order to confidently identify queries with a general navigational intent caused us to miss many queries that would typically be thought of as navigational. For example, *weather.com* and *craigslist* were missed because they had surprisingly high click entropy, although they would have been considered navigational via previous manual classifications [4, 12, 20]. In practice we observed that people used these missed queries not only to navigate to the corresponding homepage, but also to navigate to interior pages (e.g., 3.4% of all *craigslist* queries go to http://geo.craigslist.org/iso/us/ca) or related pages (e.g., 17% of all queries for *weather.com* end up at http://weather.yahoo.com). In recent years search engines have begun to support such varied uses of navigational queries by providing links directly into content of the target site. But we believe many of these queries might best be addressed by personalized navigational support.

A trend we observed among the general navigation queries highlights the potential value of looking at individual patterns of query use to identify navigational intent. Previous research suggests that the same people often issue the same queries over and over again [26, 29]. General navigation queries appeared particularly likely to be issued several times by an individual. The average number of times a person who issued a general navigation query issued the query was 2.15, which is higher than the 2.04 average number of times queries were typically issued per user. We found we could us this repeat navigational search behavior to further filter our general navigational queries to only include queries that were reused by at least some users. Nineteen of the 390 general navigation queries had a very low rate of repeat usage (less than 1.10). Several of those 19 had query strings that did not appear navigational in nature (e.g., *winning spelling bee words*). In contrast, all of the remaining 371 queries that were repeated at least sometimes by individuals were completely unambiguous. In the next section, we will investigate more deeply how we can take advantage of the fact that navigational behavior is often repeated by individuals.

4. PERSONAL NAVIGATION

Although the ability to identify queries that are used by everyone for navigation can be useful, we saw that it could be difficult to identify whether an individual intended to navigate to a particular Web resource with a particular query. We now explore how we can identify when an individual intends to navigate based on past navigational behavior. After a description of related work, we describe the algorithm we used to identify personal navigational queries, and present an analysis of the identified queries.

4.1 Related Work

Personal navigation takes advantage of the fact that individuals have long term behavior trends in the queries they issue to a

Personal Navigation Prediction Algorithm

1. Given a query q_i issued by a user,
2. Select the two most recent queries (q_{i-1} and q_{i-2}) from the user's history such that:
   - $q_{i-1} = q_i$ and $q_{i-2} = q_i$, and
   - | urls clicked(q_{i-1}) | > 0, and
   - | urls clicked(q_{i-2}) | > 0.
3. Predict the user will click $u \in$ {urls clicked(q_{i-1})} iff:
   - $q_{i-1} \neq null$ and $q_{i-2} \neq null$, and
   - | urls clicked(q_{i-1}) $\cup$ urls clicked(q_{i-2}) | = 1.

Figure 1. The personal navigation prediction algorithm. A personal navigation query is one that was used to find a particular site the past two times it was issued by the user.

search engine over time. There is significant value to understanding these long term trends [19]. For example, Wedig and Madani [30] found that the topics a user searches on are consistent over time and different from one another, and that some users repeat clicks over long time periods. In particular, we are interested in how search engines are used to return to previously viewed Web pages [7]. Adar et al. found that search engines are one of the common ways that Web pages, and particularly infrequently visited Web pages are returned to [1]. Teevan et al. [26] showed that re-finding and repeat queries were very prevalent in query logs, representing over a third of all search behavior. Sanderson and Dumais [21] examined the temporal properties of an individual's repeated searches and clicks, focusing on the aspects of repeat queries related to time. Tyler and Teevan [29] explored additional features of re-finding, such as the rank of a re-found result, the order of results clicked, the re-finding query's place in a session, the text of the result page, and the trail followed from the result page, to provide a rich picture of how elapsed time affects these features.

One interesting finding that has emerged from this research is that navigational behavior seems particularly common among repeat queries. Teevan et al. [26] looked at repeat navigational queries, which they defined as queries issued at least twice and where the same URL was clicked in the result list for each query. They found that 71% of repeated queries were navigational. Sanderson and Dumais [21] examined repeated navigational queries using the same definition, and found that around 80% of all repeat queries were navigational queries. These navigational queries were observed to be repeated over longer periods of time than non-navigational queries.

The prevalence of navigation among re-finding queries suggests that an individual's past search behavior could be very useful for predicting future navigational clicks. Significant research has gone into building search tools that use an individual's past search behavior to improve the search experience via personalization [9, 10, 23, 24, 25, 27]. For example, Teevan et al. [27] used previous clicks to indicate preferred sites for an individual to get information from, and Shen et al. [23] used previous queries to expand the user's current query. But personalization research has almost exclusively been conducted in support of finding new information, as opposed to re-finding previously viewed content. An exception is the work by Raghavan and Sever [18]. They recognized that good queries are hard to formulate, and looked at storing complex queries for future re-use. In their analysis of re-finding behavior, Teevan et al. [26] briefly explored how well

Table 2. An example of the personal navigation prediction algorithm. Personal navigation is predicted three times, once correctly and once incorrectly. When the user clicks nothing the prediction is neither wrong nor correct.

	Query	Clicked results	Predict?	Outcome
1	*wsdm*	http://wsdm2011.org	No	
2	*wsdm*	http://wsdm2011.org	No	
3	*wsdm*		Yes	Neither
4	*wsdm*	http://wsdm2011.org http://wsdm2011.org/cfp	Yes	Wrong
5	*wsdm*	http://wsdm2011.org	No	
6	*wsdm*	http://wsdm2011.org	No	
7	*wsdm*	http://wsdm2011.org	Yes	Correct

future repeat clicks could be predicted using past query behavior; in this section we expand on this past work to explore repeat personal navigational behavior.

4.2 Identifying Personal Navigation

We find that repeat navigational behavior is very useful for identifying those queries that an individual uses over and over again to navigate to the same result, even when other people do not. We say a query is a personal navigation query when the query was used to navigate to a particular Web site the past two times it was issued by a person. See Figure 1 for a detailed description of the algorithm used to identify personal navigation.

Table 2 gives an example of the algorithm in practice. Because the searcher issues the query *wsdm* at time 1 and 2 and clicks on the WSDM 2011 homepage each time, we predict the searcher will click on the WSDM 2011 homepage when they issue the query at time 3. People sometimes issue queries without clicking any subsequent result. These instances provide us no new information about the user's intent for the query, so we ignore them in calculating personal relevance. Thus, even though the user clicks on nothing at Time 3 in Table 2, we continue to predict the WSDM homepage will be clicked when we next see the query.

Occasionally, people may click on a different result from what is predicted via our personal navigation algorithm. For example, at Time 4 the searcher clicks on the WSDM Call For Papers in addition to the homepage. Because we use navigational behavior during the past two times the query was issued to predict future navigational behavior, a click on a different URL will reset the prediction. In our example, this means nothing is predicted at Time 5 or 6. At Time 7 we once again have enough history to begin predicting again.

When making the prediction we only consider instances from the user's history where the identical query string was issued. The person in our example may have interleaved other searches with the *wsdm* queries, but those other searches are ignored when predicting personal navigation for subsequent *wsdm* queries.

5. ANALYSIS

We now look more closely at the personal navigation queries we identified using the algorithm in Figure 1. After an overview of what personal navigation queries look like, we show that they allow us to identify with high accuracy navigational intent for many queries. We then dive more deeply into at how consistent personal navigation behavior is across time, and explore individual differences and group behavior.

Table 3. The personal navigation results identified for the query *lottery*, along with a count of how often the query is used to navigate to the associated URL.

URL	Count	Percent
http://www.lottery.com/	85	41.26%
http://www.michigan.gov/lottery	35	16.99%
http://www.illinoislottery.com/	22	10.68%
http://www.calottery.com/default.htm	11	5.34%
http://www.palottery.state.pa.us/	11	5.34%
http://www.masslottery.com/	10	4.85%
http://national-lottery.co.uk/	7	3.40%
http://www.nylottery.org/	5	2.43%
http://www.txlottery.org/	4	1.94%
http://www.georgialottery.com/	3	1.46%
http://www.valottery.com/	3	1.46%
http://www.kylottery.com/	2	0.97%
http://www.mdlottery.com/	2	0.97%
http://www.nc-educationlottery.org/	2	0.97%
http://www.oregonlottery.org/	1	0.49%
http://www.tnlottery.com/	1	0.49%
http://www.txlottery.org/export/...	1	0.49%
http://www.walottery.com/	1	0.49%

5.1 Overview

In this section we give a general overview of what personal navigation queries look like. Table 3 shows an example of the different sites identified as the target of personal navigation for different individuals for the query *lottery*. The most common personal navigational target following the query was http://www.lottery.com/, but over half of the instances show the query being used to navigate to other lottery websites, including the homepages for the Michigan lottery, the Illinois lottery, and the California lottery. As another example, the query *enquirer* was also used by some people for personal navigation. Although it was used to navigate to the National Enquirer Web site (http://www.nationalenquirer.com/) 34 times, it triggered for Cincinnati's newspaper homepage (http://enquirer.com/) 33 times, and to the news section within that site 10 times.

As can be seen in the above examples, it is not always obvious what the intended navigational target is merely from the query string. Further, in many cases the query does not even appear navigational. For example, although many searchers use the query *bed bugs* to learn general information about the insect, one user in our sample used it to navigate repeatedly to the URL http://www.medicinenet.com/bed_bugs/article.htm. Most likely this is a result that was originally found via an initial informational search, and then later returned to in a navigational manner using the same query, in a manner similar to the query chains observed by Tyler and Teevan [29].

Table 1 presents the same statistics we looked at for general navigation queries for personal navigation queries, based on the personal navigation queries identified from May 8 to May 9 when using one month of search history data (see Column 2 of Table 4, to be discussed in greater detail later). General navigation queries identified via the personal navigation approach were excluded from the analysis. Unlike for general navigation, where everyone issuing the query is assumed to have a navigational intent, for personal navigation only the subset of individuals using the query for navigation are assumed to have that intent. Thus the personal navigation column of Table 1 only reports the behavior of people that we have identified as using the query for navigation.

In Table 1 we see that personal navigation queries and how they were used is quite different from general navigation queries and the typical Web search engine query. The length of personal navigation queries falls between the two other types, being shorter than the average query, at 1.84 words per query and 11.73 characters, but are not as short as the very common general navigation queries (1.46 words, 9.00 characters). While brevity is advantageous for navigational queries, the personal navigation queries were typically less popular than the very popular general navigation queries, and thus may have required additional text to fully specify what was being sought. Interestingly, personal navigation queries were less likely than both general navigation queries and queries in general to contain a URL fragment.

Given the differences in how personal and general navigation queries are identified, it is not surprising that personal navigation queries were much less popular than general navigation queries, occurring at an average frequency of 245,992, or less than a tenth of the time that general navigation queries did. The fact that the frequency is even lower than the popularity of the average Web search queries reflects the fact that the personal navigation frequency number only represents instances where the queries are identified as being used for navigation. The queries could have been used by other users for other purposes, and when all uses are considered the average popularity of the personal navigation queries falls between the other two types, at 764,323. Still, only a small fraction (< 0.06%) of the unique personal navigation query instances occurred at even a tenth of the frequency as our general navigation queries (i.e., more than 1000 times). But in terms of query volume, these queries accounted for about half (60.0%) of the personal navigational query volume.

Personal navigation queries were often used uniquely by just one person. For example, there were a number of instances where the name of a person who is not famous was repeatedly used by one individual to navigate to the named person's homepage. The earlier *bed bugs* query is another example of an uncommon personal navigation query (although in this case the query is common). A large majority (97.7%) of the unique personal navigational query instances were issued 25 times or fewer, and these account for 21.7% of the personal navigation query volume.

While many personal navigation queries were not issued by many different people, they did tend to be issued again and again by the same person. Each personal navigation query was issued on average 4.59 times per user who issued them, which is much higher than what we observed for queries as a whole (which were issued 2.04 times per user) and for general navigation queries (which were issued 2.15 times per user). This high repeat rate reflects, in part, the fact that we identify personal navigation queries based on repeat usage. To even trigger as a personal navigation query, the query must have already been used at least twice by the same person.

The queries that were used by individuals for personal navigation were often used by other people in other ways. The click entropy across all people who issue the personal navigation queries, regardless of intent, averaged 1.41, which is much higher than the average click entropy across all people who issue the general navigation queries (0.66). Not all of the variation seen within personal navigation queries came from some people having non-navigation intents for the same query. When only instances of

Table 4. The coverage and accuracy of the personal navigation prediction under various different conditions. The first three columns show the performance for three different aggregation and test periods. Even with only a week of history, personal navigation achieves high coverage and accuracy. The next two columns show performance based on data collected days or weeks prior the current query, revealing that a person's navigation behavior is fairly consistent over time. The last column shows that predictions based on people grouped by location can achieve higher coverage at the expense of accuracy.

		Personal			Over Time		Group
		Week	Month	Month	Offline	Over time	Online
Aggregation period length		1 week	1 month	1 month	1 week	1 week	1 month
Aggregation period start		May 7, 2010	April 7, 2010	May 7, 2010	May 5, 2010	May 5, 2010	May 7, 2010
Test period length		2 days	2 days	2 days	2 days	2 days	2 days
Test period start		May 15, 2010	May 8, 2010	June 8, 2010	May 12, 2010	June 8, 2010	June 8, 2010
Training type		Online	Online	Online	Offline	Offline	Online
Total test queries with clicks		52,105,793	51,046,291	83,469,052	83,012,992	83,469,052	83,469,052
All queries	Number of predictions	6,755,781	7,653,798	12,676,081	9,147,937	4,147,013	23,672,763
	Good predictions	6,324,005	7,153,234	12,026,592	8,674,570	3,733,728	18,881,413
	Bad predictions	431,776	500,564	649,489	473,367	413,285	4,791,350
	Coverage	12.97%	14.99%	15.24%	11.02%	4.97%	28.36%
	Accuracy	93.61%	93.46%	94.88%	94.83%	90.03%	79.76%
Excluding general navigation	Number of predictions	3,262,709	3,816,455	6,903,972	4,894,927	2,206,636	15,191,466
	Good predictions	2,954,494	3,503,734	6,425,328	4,582,165	1,935,422	11,851,930
	Bad predictions	308,215	312,721	478,644	312,762	271,214	3,339,536
	Coverage	7.43%	8.89%	9.64%	6.84%	3.08%	21.22%
	Accuracy	90.55%	91.81%	93.11%	93.61%	87.71%	78.02%

navigational intent are considered, we find that the click entropy is still non-zero, at 0.37. This means that different people were using the same personal navigation queries to get to different URLs. We will look more closely at this in Section 5.4.

Although we saw earlier that the target of general navigation was almost always ranked first, we find that for personal navigation the target was more likely to sometimes be ranked lower in the result list. The average rank of the personal navigation target was 1.31, compared with 1.03 for the general navigation target. Similarly, the personal navigation result was not first 84.55% of the time, compared with 98.16% of the time for general navigation. And not only were personal navigation targets ranked lower than the general navigation targets, but they were also ranked lower than the results found for Web queries overall (the average click rank for Web queries was 1.17, with people clicking on the first result 91.38% of the time). This suggests that there is a larger opportunity to support personal navigation via re-ranking than there is to support general navigation, and perhaps even than there is for improving overall Web search ranking.

5.2 Coverage and Accuracy

We can assess the value of personal navigation by looking at how often personal navigation triggered (i.e., the *coverage* of personal navigation) and how often the personal navigation algorithm correctly predicted what the user would click given it triggered (i.e., the *accuracy* of the personal navigation prediction). We calculate coverage and accuracy as follows:

Coverage: The total number of personal navigation predictions made divided by the total number of queries issued. Note that coverage does not count the first two times a personal navigation query was issued by an individual as predictions, since no prediction is made for those queries.

Accuracy: The number of correct predictions made divided by the total number of correct or wrong predictions made. Instances where a prediction is made but the user does not click on a result are ignored in the calculation of accuracy, because in these cases we do not know if the prediction was correct or not. Instances where the user clicks on the predicted result but also clicks on one or more other results are considered incorrect predictions.

To calculate the coverage and accuracy of our personal navigation approach, we began by aggregating historical user data over a period of time to make the initial personal navigation predictions.

Aggregation period: The period of user history used to initially identify personal navigation behavior.

Aggregation start: The day the aggregation period began.

We then tested the personal navigation predictions over a subsequent period of time (i.e., the *test period*), updating the predictions dynamically so that the last two times the individual issued a query were always what were used to make the prediction. For example, if an individual issued the query *bed bugs* once in the aggregation period, no prediction was made the first time the query was issued in the test period, but a prediction was made the second time. Since predictions are always made based on the two most recent issuances (with clicks) of the query, once a bad prediction is made, no prediction is made for the query for at least the next two times the query is seen.

Test period: The length of time during which personal navigation predictions were made.

Test start: The day the test period began.

The first three columns of Table 4 show the coverage and accuracy for three different aggregation and test periods. The first column represents one week of aggregation, and the next two

represent one month of aggregation. In all three cases a two day test period is used. As can be seen, a significant portion of search engine traffic can be easily identified as personal navigation. With only a week of aggregation data, we find it is possible to predict what a person will click for 12.97% of all queries. With a month of aggregation data, this number increases to 14.99% to 15.24% of all queries. We suspect that there would be an even larger increase in coverage if the search engine had access to a more reliable and long-lived way to identify users than cookies. The accuracy of the personal navigation predictions was very high, ranging from 93.46% to 94.88%. This number would be even higher if queries for which the predicted target was clicked and other results were clicked were considered correct.

5.2.1 Compared with General Navigation

These coverage and accuracy numbers are significantly higher than what we observed for the general navigation queries we identified. General navigation queries represented about 12.02% of all queries with clicks, and accurately predicted the user's click following the query only 71.74% of the time. Of course, some of the queries identified via our personal navigation approach were also general navigation queries, as we saw that individuals tended to repeat general navigation queries. Because these queries were ones for which the target was considered to be already known even without an individual's query history, we also looked at the coverage and accuracy of personal navigation when general navigation queries were excluded. In this case, we see the coverage is 7.43% with one week of aggregation data, and 8.89% to 9.64% with one month of aggregation data. This means the general navigation queries accounted for 5.5% to 6.1% of the personal navigation volume, or that about half of the general navigation coverage of 12.02%. If a search engine were to identify both general navigation and personal navigation according to the approaches described here, it could accurately predict navigational intent for over 21% of all queries.

The accuracy for personal navigation falls by just over a point when general navigation is excluded, to 91.18% to 93.11%. Overall, the accuracy for the personal navigation predictions is much higher than for the general navigation predictions (71.74%). However, when general navigation queries trigger as personal navigation, the prediction is over 20% more accurate than when predicted based on aggregate behavior alone.

An advantage to our approach to identifying general navigation queries over personal navigation queries is that users can reap the benefit of the prediction immediately, the first time the query is issued. In contrast, personal navigation requires the query to be issued several times. It may be possible to make earlier and more accurate navigational predictions using less history by combining the click pattern trends of search engine users in general with an individual's own search history. For example, a search engine could predict navigational intent with only one query of history for queries that appear likely to be used by most people to get to the same URL, but that do not pass the general navigation threshold. In the next two sections we explore how temporal patterns can impact the prediction quality, and look at the value of using the browsing behavior of subsets of search engine users to predict navigation queries.

5.3 Consistency of Predictions over Time

Until now, we have looked at predicting personal navigation intent for a query based on the individual's two most recent issuances of the query, and not taken into account the elapsed time or previous issuances of the query. In this section, we turn to the consistency of the personal navigation predictions over time. We begin by showing that personal navigation intent can be predicted with data collected days or weeks prior the current query. We then show that the more a person uses a personal navigation query, the more confident we can be in our prediction for that query, and we discuss some of the challenges with supporting this consistent behavior in the face of search result dynamics.

5.3.1 Offline Predictions

The personal navigation prediction algorithm described in Section 4.2 assumes that the search engine has instant access to a person's search history, and is able to use the two most recent issuances of a query to make a prediction. However, instant and fast access to an individual's query history can be expensive. It may be beneficial to calculate the predictions in advance (i.e., *offline*) rather than in real time (i.e., *online*), and store the predictions in a static lookup table for quick and easy access.

Offline: If the last two clicked query instances in the aggregation period resulted in the same single click, these two instances are used to predict what will be clicked in the test period.

Online: Uses the most recent past two clicked query instances prior to prediction (from aggregation period or test period) to predict click in test period.

The coverage and accuracy numbers we have discussed so far for personal navigation have been based on online prediction. The value of offline predictions can be seen in the section of Table 4 marked "Over Time." We find that there is still significant coverage and accuracy when personal navigation is computed offline. Using one week of aggregation data and testing offline on two days of test data, personal navigation intent can be predicted for 11.02% of all queries with 94.83% accuracy. This represents only about a 2% drop in coverage from the online prediction with a week of aggregation, and accuracy is not hurt at all.

To explore whether personal navigation predictions can extend over even longer periods of time, we looked at using a training period situated about one month before the test period, and calculated the coverage and accuracy during the test period using the month-old offline predictions. We found the coverage was 4.97% and the accuracy was 90%, both of which are still fairly high. This means that in many cases, the personal navigation queries people use are ones they have been using for weeks. Given the longevity of such queries in the system, search engines may be able to devote additional resources to support these queries, even if such things can only be done slowly. For example, only 64% of the URLs found via personal navigation currently have deep links available from the Bing service. However, the search engine could crawl the remaining 36% to provide deep links for those results as well. And by storing a little more context per user, the search engine could notify the user of updated content on the target URL since the user's last visit.

The fact that personal navigation queries can be predicted based on data collected well in advance of the prediction is consistent with previous research that has shown that unlike many other types of repeat queries, repeat navigational queries extend over periods of time. For example, Sanderson and Dumais [21] found that navigational queries are less likely to be repeated by users within a few days than queries with a more information seeking focus, and navigational queries are more likely to be repeated at later points in time. This may be because search engines are disproportionately useful when users want to return to infrequently revisited sites [1, 6].
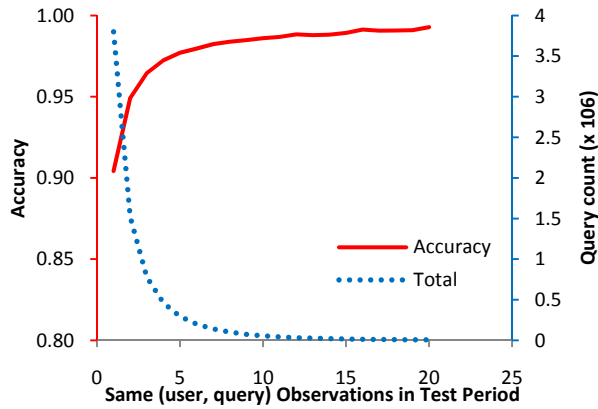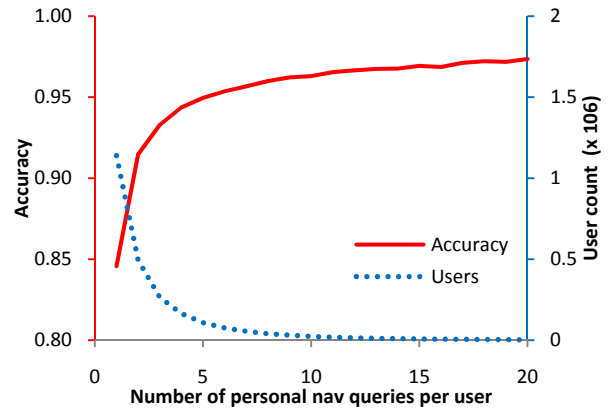
Figure 2. The accuracy of the personal navigation prediction as a function of how often the individual has used the same query for personal navigation.

Figure 3. The accuracy of the personal navigation prediction as a function of how often a user triggers a personal navigation for any query.

5.3.2 Improving Predictions over Time

We also find that our prediction of personal navigation intent got better as we saw more data from the same individual for the same query over time. This can be seen graphically in Figure 2. The solid line represents the accuracy of the prediction based on how many times the individual has issued the predicted query in the past. When a person had only issued the query once before, the prediction accuracy is 90.4%. This quickly rises to almost 100% accuracy when the individual has issued the same query and clicked the same result ten or more times.

The tradeoff, of course, to requiring many past examples of navigation to predict future navigation is that this reduces the coverage. Also shown in Figure 2 (dotted line) is the number of predictions possible based on how many times the individual had issued the predicted query in the past. There were many more predicted instances when the query was issued only a few times than when it was issued ten or more times. As suggested earlier, it may be possible to couple general navigation type features with personal features to improve the accuracy earlier on, and we explore using group data to identify navigation in the next section.

We also explored whether some people were more likely to successfully use navigational queries than others. We found that while most people used navigational queries at some point, we could predict the behavior of people who used them frequently better than we can predict the behavior of people who used them rarely. There were relatively few personal navigation super users, meaning most people used personal navigation on occasion, and few people used it over and over again. Almost half (45.9%) of the people who triggered a personal navigation query did so only once, and 99.1% triggered personal navigation 10 times or fewer, although this probably understates the true use of personal navigations per person due to churn in the cookie-based user IDs. Just as we saw that individuals who navigate with the same query over and over again were more likely to be predictable in their navigation, so too we find that people who issued personal navigation queries many times were more likely to do so in a predictable manner, regardless of whether the particular personal navigation query was the same in all instances or not. As can be seen in Figure 3, people who triggered personal navigation only once had an accuracy of 84.5%, and twice of 91.5%. This moved towards an asymptote of around 97.4%.

5.3.3 Challenges with Time

There is an interesting challenge that arises from the fact that personal navigation queries were used consistently over time, stemming from the fact that search results change rapidly over time [22]. Although people develop expectations about repeat search results based on the results they have seen before [26], the navigational target a person expects can disappear from the result list for the associated query over the course of weeks. The prevalence of personal navigation behavior that we observed is likely to be an underestimate of the true desire for personal navigation for this reason; when the link no longer occurs in the list, we cannot tell from the log data if the searcher intended to visit it. Additionally, previous viewed results that remain in the top ten but that change rank to appear lower than they did when initially encountered are probably often missed by the user even though they are shown [13]. Support for personal navigation could enable search engines to preserve previous interactions even in the face of such changes.

5.4 Individual and Group Navigation

We now look more closely at the differences – and similarities – between the different people who used the same queries for personal navigation. We begin by exploring the diverse navigational ways that the same query was used. We then show that although different people often used the same query to navigate to different results, there were also consistencies across groups of people that we can take advantage when identifying navigational queries.

5.4.1 Personal Navigation Varies by Individual

In Section 3.2 we identified queries that were popular and generally used for navigation by all of the people who issued them. In this section we show that the personal navigation queries we identified are not just queries on the fringes of having been identified as general navigation, but instead are queries that just happen to be used consistently by the individual in question.

Many personal navigation queries were used similarly by all of the people who issued them and had very low click entropy overall; 69.54% of the query instances had a general click entropy of less than 1.0, the same criteria we used to identify general navigation. The first row of Table 5 gives some examples of personal navigation queries with low click entropy. However,

most low click entropy queries had low click entropy because they were used by just one or two people. Of the unique personal navigation queries with a click entropy of less than 0.5, 87.70% were issued by five people or fewer. For these queries, there is nowhere near enough data to generalize the individual's behavior to other people. As can be seen in Table 5, while some queries with low click entropy used by only a few people appear to clearly refer to a specific resource (e.g., *azflyfishing* or *www.chili.org*), not all do (e.g., *lesson on golf* or *media gossip*). In contrast, the queries used for personal navigation by many people (e.g., *qvc*) are heavily skewed towards referencing particular resources, and are probably borderline general navigation queries.

Many other personal navigation queries were used differently by the people who issue them. Examples of personal navigation queries with high click entropy can be seen in the bottom row of Table 5. Unlike the queries with low click entropy, these queries most often do not have obvious resources associated with them. For 23.72% of the personal navigation queries that were issued more than 10 times, the same query was used to navigate to more than one resource. An example of this is the *lottery* query shown in Table 3; people use it to navigate to results other than http://www.lottery.com 58.74% of the time. Even seemingly unambiguous queries are used by different people to navigate to different locations. For example, on the surface it appears obvious that the query *real estate.com* is intended to navigate to the site http://www.realestate.com. However, for only five of the 23 times that query is used for personal navigation does the query lead to a click on the obvious target. Instead, it is much more likely to be used to navigate to http://realestate.msn.com or http://www.realtor.com.

5.4.2 *Personal Navigation Consistent across Groups*

As we have discussed, there are a number of popular queries used for personal navigation with relatively low click entropy. Some of these queries may warrant special treatment for many people, or for specific groups of people, even when we are lacking significant personal history for those people. For this reason, we explored identifying personal navigation behavior across groups. The goal here was to combine general and personal navigation behavior to provide personal navigation support with less personal history than required in our initial algorithm.

To explore this, we extended our approach for identifying personal navigation queries in the most straight forward manner possible: instead of using a person's user ID to identify relevant queries from that person's history, we used the person's location to identify relevant queries from the location's history. Thus the two most recent issuances of the query from the location were used to make the prediction. The person issuing the query need not have ever issued the query before.

The results to this approach are shown in the last column of Table 4. As expected, we make predictions for many more queries in this manner, for a 28.4% coverage. A number of these queries were general navigational queries, but 21.22% were not. This is an 11.58 point improvement above using an individual's history. This improvement comes at a cost, however. The accuracy of the prediction drops significantly, to 78.0%. In part, this may be due to our naïve implementation of group navigation. It may be possible to see greater accuracy improvements with little cost to coverage by using looking at more queries than just the two most recent in time. In any case, it seems likely that combining individual query history with group information and information about Web searchers in general will enable increased coverage for our predictions while maintaining a very high accuracy.

Table 5. Several examples of queries with high and low click entropy, as relates to the number of people who use the query for personal navigation.

		Personal navigation use	
		Few people	**Many people**
Click entropy	Low	azflyfishing great bay volleyball lesson on golf media gossip new moon bar stanford lane library www chili org	drudge report eharmony login msn games myyearbook pandora photobucket qvc
	High	biggest spider cooking videos firewood rock and gem tiny house plans url shortener wholesale clothes	american idol flash games hidden object games horoscopes justin bieber kate gosselin tiger woods

6. CONCLUSION AND FUTURE WORK

In this paper, we have demonstrated that there is a rich opportunity for straightforward search result personalization in support of personal navigation. We showed that it is possible to identify general navigation queries and to thus know what people will click on for many queries a search engine sees, but that using an individual's behavior to identify personal navigation allows for greater coverage and higher accuracy. We demonstrated that while personal navigation queries share some similarities with general navigation queries (e.g., they are relatively short), they tend to look more like other search engine traffic as a whole. Queries that are typically informational become navigational on a personal level. Personal navigation queries are relatively unpopular, often being issued by only a few people, and click behavior revealed that personal navigation queries are often used by different people in different ways, including to navigate to different search results.

The patterns we observed in the consistent use of personal navigation over time and across locally proximal groups of people suggest ways our ability to predict personal navigational intent might be improved. For example, we saw that prediction accuracy improved significantly for searchers who issued the same personal navigation query many times or who issued multiple personal navigation queries. There may be additional contextual information that could be used as well to improve accuracy. However, the more specific the information needed to make the prediction is, the lower the coverage. We believe it will be possible to improve coverage by coupling this additional personal information with group and aggregate data, allowing for the capture of both increased coverage and increased accuracy.

Another potential improvement would be to allow people to sometimes make mistakes in what they click following a personal navigation query. Right now, if the searcher clicks on another result in addition to what is predicted, the algorithm stops predicting what will be clicked for that query for at least the next two times the query is issued by that user. But if the person has a long history of issuing that query to navigate to a particular result, it may make more sense to continue to make the prediction than to immediately try to correct it.

There are many opportunities for search engines to use what is presented in this paper to identify navigational queries and their targets, and use that information provide significant benefit to the user. The navigational target, which, in the case of personal navigation is more likely not to be displayed first than for other types of queries, can be displayed first. And in cases where the result no longer appears in the result list, it can be added into the list. The interface can be designed to support the desired intent by providing, for example, links directly into the site's content [8] or access to appropriate meta-data or site functionality. Additionally, search engines can use information operationally, to provide better caching for navigational queries [17] or provide the user with more appropriate advertisements [2].

Looking beyond search, we believe there is an opportunity to apply the same personalization approach presented here to other activities in which people have strong patterns of behavior. For example, Web browsing behavior is remarkably consistent, and in many cases it is easy to predict the next page an individual will visit given their browsing history and the page they are currently on. A Web browser could use this information to personalize the browsing experience by, for example, pre-fetching the next page for faster rendering. Similarly, application use is often very consistent within an individual, with the same person regularly performing the same pattern of activities.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Adar, E., J. Teevan, and S. T. Dumais. Large scale analysis of Web revisitation patterns. In *Proceedings of CHI '08*, 2008, 1197-1206.

[2] Ashkan, A., Clark, C., Agichtein, E., and Guo, Q. Estimating ad click-through rate through query intent analysis. In *Proceedings of Web Intelligence & Intelligent Agent,* 2009, 222-229.

[3] Ashkan, A., Clarke, C., Agichtein, E., and Guo, Q.. Classifying and characterizing query intent. In *Proceedings of ECIR '09*, 2009, 578-586.

[4] Baeza-Yates, R., Calderón-Benavides, L., and González, C. The intention behind Web queries. In *Proceedings of SPIRE '06*, 2006, 98-109.

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

[6] Bruce, H., Jones, W. and Dumais, S. Keeping and re-finding information on the Web: What do people do and what do they need? In *Proceedings of ASIST '04*, 2004, 1-10.

[7] Capra, R. and Pérez-Quiñones, M. A. Using Web search engines to find and refind information. *IEEE Computer*, 38 (10), 2005, 36-42.

[8] Chakrabarti, D., Kumar, R. and Punera, K. Quicklink selection for navigational query results. In *Proceedings of WWW '09*, 2009, 391-400.

[9] Chirita, P., Firan, C., and Nejdl, W. Summarizing local context to personalize global Web search. In *Proceedings of SIGIR '06*, 2006, 287-296.

[10] Dou, Z., Song, R., and Wen, J. R. A large-scale evaluation and analysis of personalized search strategies. In *Proceedings of WWW '07*, 2007, 581-590.

[11] Herrera, M. R., de Moura, E. S., Cristo, M., Silva, T. P., and da Silva, A. S. Exploring features for the automatic identification of user goals in Web search. *IP&M*, 46(2): 2010, 131-142.

[12] Jansen, B. J., Booth, D., and Spink, A. Determining the informational, navigational, and transactional intent of Web queries. *IP&M*, 44(3): 2008, 1251-1266.

[13] Joachims, T., Granka, L., Pan, B., Hembrooke, H., and Gay, G. Accurately interpreting clickthrough data as implicit feedback. In *Proceedings of SIGIR '05*, 2005, 154-161.

[14] Kang, I.-H. and Kim, G. Query type classification for Web document Retrieval. In *Proceedings of SIGIR '03*, 2003, 64-71.

[15] Lee, U., Liu, Z, and Cho, J. Automatic identification of user goals in Web search. In *Proceedings of WWW'05*, 2005, 391-400.

[16] Lu, Y., Peng, F., Li, X., and Ahmed, M. Coupling feature selection and machine learning methods for navigational query identification. In *Proceedings of CIKM '06*, 2006, 682-689.

[17] Ozcan, R., Altingovde, I. S., and Ulusoy, Ö. Utilization of navigational queries for result presentation and caching in search engines. In *Proceedings of CIKM '08*, 2008, 1499-1500.

[18] Raghavan, V. and Server, H. On the reuse of past optimal queries. In *Proceedings of SIGIR '95*, 1995, 344-350.

[19] Richardson, M. 2008. Learning about the world through long-term query logs. *TWEB,* 2 (4), 2008, 1-27.

[20] Rose, D. E. and Levinson, D. Understanding user goals in Web search. In *Proceedings of WWW '04*, 2004, 13-19.

[21] Sanderson, M. and Dumais, S. Examining repetition in user search behavior. In *Proceedings of ECIR '07*, 2007, 597-604.

[22] Selberg, E. and Etzioni, O. On the instability of Web search engines. In *Proceedings of RIAO '00*, 2000, 223-236.

[23] Shen, X., Tan, B. and Zhai, C. X. Implicit user modeling for personalized search. In *Proceedings of CIKM '05*, 2005, 824–831.

[24] Speretta, M. and Gauch, S. Personalized search based on user search histories. In *Proceedings of WI '05*, 2005, 622-628.

[25] Sugiyama, K., Hatano, K., and Yoshikawa, M. Adaptive Web search based on user profile constructed without any effort from user. In *Proceedings of WWW '04*, 2004, 675–684.

[26] Teevan, J., Adar, E., Jones, R. and Potts, M. Information re-retrieval: Repeat queries in Yahoo's logs. In *Proceedings of SIGIR '07*, 2007, 151-158.

[27] Teevan, J., Dumais, S. T., and Horvitz, E. 2005. Personalizing search via automated analysis of interests and activities. In *Proceedings of SIGIR '05*, 2005, 449-456.

[28] Teevan, J., Dumais, S. T., and Liebling, D. J. 2008. To personalize or not to personalize: Modeling queries with variation in user intent. In *Proceedings of SIGIR '08*, 2008, 163-170.

[29] Tyler, S. and Teevan, J. 2010. Large scale query log analysis of re-finding. In *Proceedings of WSDM '10*, 2010, 191-200.

[30] Wedig S. and Madani, O. A large-scale analysis of query logs for assessing personalization opportunities. In *Proceedings of KDD '06*, 2006, 742-747.