

Information Re-Retrieval: Repeat Queries in Yahoo's Logs

Jaime Teevan
MIT, CSAIL
Cambridge, MA 02138 USA
teevan@csail.mit.edu

Eytan Adar
University of Washington, CSE
Seattle, WA 98195 USA
eadar@u.washington.edu

Rosie Jones
Yahoo! Research
Burbank, CA 91504 USA
jonesr@yahoo-inc.com

Michael A. S. Potts
Yahoo! Research
Sunnyvale, CA 94089 USA
mpotts@yahoo-inc.com

ABSTRACT

People often repeat Web searches, both to find new information on topics they have previously explored and to re-find information they have seen in the past. The query associated with a repeat search may differ from the initial query but can nonetheless lead to clicks on the same results. This paper explores repeat search behavior through the analysis of a one-year Web query log of 114 anonymous users and a separate controlled survey of an additional 119 volunteers. Our study demonstrates that as many as 40% of all queries are re-finding queries. Re-finding appears to be an important behavior for search engines to explicitly support, and we explore how this can be done. We demonstrate that changes to search engine results can hinder re-finding, and provide a way to automatically detect repeat searches and predict repeat clicks.

Categories and Subject Descriptors

H3.3 [Information storage and retrieval]: Search and retrieval

General Terms

Measurement, Experimentation, Human Factors.

Keywords

Query log analysis, Web search, re-finding, repeat queries.

1. INTRODUCTION

Thanks to the ubiquity of the Internet search engine box, users have come to depend on Web search engines both to find new information and to re-find previously viewed information. A recent Pew Internet and American Life report showed that Internet searches are a top Internet activity, second only to email [16]; in a study of Web users [9], 17% of those surveyed reported "Not being able to return to a page I once visited," as one of "the biggest problems in using the Web." The effect of this is that knowledge workers are estimated to waste 15% of their time because they cannot find information that they know already exists [7]. Despite these known problems, the use of keyword search engines for re-finding has not been significantly studied.

While many searches are for new information, a significant use of search engines is to find information that was found before. For example, a query or keyword is often used to "bookmark" a Web page. In this paper, we build on earlier work [21] to explore how keyword search is used for re-finding. We analyze the queries and result clicks of 114 anonymous Yahoo users over the course

of a year. Our analysis demonstrates that re-finding queries are common and provides a detailed characterization of them. Given the pervasiveness of re-finding queries, we explore which search engine features support or hinder re-finding. In particular, we concentrate on changes in rank and demonstrate the detrimental impact of rank changes on this type of task. Making use of our understanding of re-finding behavior, we describe algorithmic methods to detect re-finding intent and suggest ways in which search engines can better support this behavior.

Log studies like the one presented here are valuable because they give a large-scale, realistic picture of users' actions. However, they give no insight into underlying motivation. To study re-finding through log analysis, it was necessary to try to glean from the data those queries which were intended to re-find information rather than find new information. Re-finding intent was approximated by looking for repeated clicks on the same search result in response to queries issued by the same user at different times. The query used to re-find the result may or may not be the same as the query used to find it originally. For example, if a person searched with the query "KPCC Southern California Public Radio" and clicked on the result <http://www.scpr.org>, and then later clicked on the same result while searching for "spcr", the behavior was considered re-finding. Because of our limited ability to automatically distinguish re-finding from finding behavior in the query logs, our observations were supplemented with a separate additional controlled experiment of a panel of 119 volunteers where a re-finding task was explicitly defined.

No matter how re-finding is approximated in the logs, analysis reveals it is very common. Forty percent of all observed queries led to a click on a result that was also clicked during another query session by the same user, and nearly 30% of all URLs clicked in the data set were clicked by the same user more than once. As we will demonstrate, the impact of the interplay between this common behavior and changing result rankings has a cost in terms of session time. As a way of dealing with this problem, we discuss simple but effective ways to automatically detect re-finding queries and implications for search engine design.

2. RELATED WORK

Re-finding behavior has recently attracted considerable interest [1, 3, 4, 6, 20, 21, 22]. Many re-finding studies have been limited to small-scale laboratory or interview based studies. Such studies of re-finding have consistently found that people tend to rely on meta-data about their target [6], for example, re-finding previously viewed content via known paths [4, 22]. Thus if someone originally encountered a piece of information via a search engine, that person is likely to try to repeat the same query to find that same information again. However, because people process encountered information to varying degrees, some re-finding may rely heavily on meta-data learned during the initial

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGIR '07, July 23–27, 2007, Amsterdam, The Netherlands.

Copyright 2007 ACM 978-1-59593-597-7/07/0007...\$5.00.

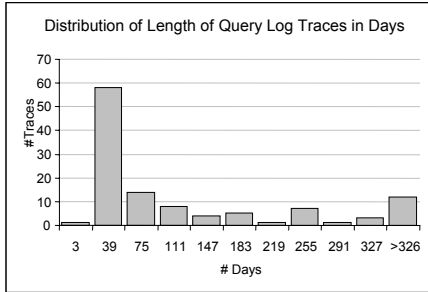


Figure 1: Distribution of length of search traces in days (average: 97 days, median: 37 days)

behavior than laboratory and observational studies, and gives a very realistic picture of people’s actions, although it gives no insight into people’s underlying motivation. Log analysis has shown that Web site re-visitation is very common [5, 14, 19], with estimates of the portion of Web page visits that are re-visits reaching up to 80% [5]. While many of these re-visitations occur shortly after the first visit (*e.g.*, during the same session using the back button), a significant number occur after a considerable amount of time has elapsed. The results of Web site re-visitation studies have informed Web browser history and back button functionality.

While large-scale studies have been done on query logs (*e.g.*, [10]), surprisingly, there has been little analysis of re-visitation and re-finding. These studies have found that most queries are issued only once or twice. These results are particularly interesting when considered in light of our study, as we see that individuals are very likely to repeat queries. It is likely that many of the repeat queries seen by a search engine are repeated by the same individual(s).

Some log analysis studies have looked at queries clustered by topic [2, 17, 24]. Several studies have investigated queries in aggregate over time, to understand changes in popularity [23] and uniqueness of query topics at different times of day [2]. Wedig and Madani showed that topics for a user are consistent over time and different from one another, and that some users repeat clicks over long time periods [24]. Others have analyzed queries over time for individuals, but focused on short periods of time such as query sessions [11, 13].

Sanderson and Dumais [18] confirmed re-finding behavior observed by us in previous work [21] and extended that work by examining temporal properties of repeated searches and clicks over a period of 90 days. They focused on the temporal aspects of repeat queries, finding, for example, that navigational queries are repeated over longer periods of time than non-navigational queries. Our work is unique in that it looks at combinations of query and click patterns for anonymous individuals over long time periods (one year). Because of the long time period studied we are able to characterize the different ways users express the repeat query intent and explore how they deal with result list changes.

3. STUDY METHODOLOGY

We analyzed search traces of queries issued to the Yahoo search engine over a period of 365 days (August 1, 2004 to July 31, 2005) by 114 users. Search traces were considered for inclusion in the study if they included queries issued during at least four of

encounter, while some may look very similar to searches for new information [22]. We observe repeat query behavior in the logs that seems to represent the entire spectrum.

Log analysis allows researchers to observe a greater variety of

the last ten days of the sample period. The average trace was 97 days long (see Figure 1). The study was conducted in accordance with Yahoo’s terms of service and privacy policy. All traces analyzed were strictly anonymous; data was never used to match a search trace with an identity. Furthermore, results reported in this paper rely solely on aggregated statistics, and examples are purely illustrative.

For the analysis described here, we focused primarily on the large majority of queries for which there was a click on a result page; we excluded next-page clicks, clicks on alternate query suggestions and instances where there was no click at all. The data were not filtered to remove search spam or robot/mechanical searches. Some of the strongest repeated-search repeated-click traffic may come from robots and how those may be detected based on re-finding behavior is briefly discussed later.

In the analysis, we were not interested in very short-term query repetitions. Short-term repeat queries were most likely a result of page refreshes or back-button clicks [14]. To remove short-term repeat queries from our data, we considered all instances of a query that occurred within thirty minutes of an identical query to be a single query. The threshold was chosen because there was a clear inflection in the data between the frequency at which searches were repeated before and after this point, suggesting that the observed behavior was different.

We looked at the following information in our analysis: The query terms issued; an anonymous key distinguishing the user; the time the query was issued; what results were clicked and when; and their position in the result list. In total, 13,060 queries were observed (an average of 115 per trace). The average query length (2.7 words), and average number of results clicked per result page was similar to what is reported elsewhere [10].

4. IDENTIFYING REPEAT QUERIES

To successfully identify repeat queries in this data, it was necessary to associate queries by inferring the intent of the user, rather than relying on the exact query string being repeated. Many users repeated past query strings perfectly (*e.g.*, “bbc world service”). Of the 13,060 query instances, 4,256, or 33%, were an exact repeat of a query issued by the same anonymous user ID at another time. In contrast, only 860, or 7%, of all queries were issued by more than one user.

Often, when identical queries occurred in the same trace, the user associated with the trace clicked on the same results following the identical query issuances. We also found a number of identical clicks that occurred following different queries (*e.g.*, “pennsylvania lottery” and “pennsylvania powerball”). However, even identical queries did not guarantee a repeat click; it was quite common for repeat queries to lead to unique clicks. This section proposes a taxonomy of repeat queries, based on various combinations of query and click comparisons, and discusses their probable underlying intent. In defining the taxonomy we are interested in both the query issued and the set of clicked results. Table 1 represents all possible classifications based on these two dimensions (query string and click-through sets).

A number of the classes shown in Table 1 are very uncommon (*e.g.*, queries for which there are multiple identical clicks). In this paper, we concentrate our efforts on understanding the popular categories that are likely to include re-finding intent (bolded in Table 1). The broader the class of query captured by the category, the more likely it is to include both re-finding intent and

Table 1. A classification of different query types.

All queries: 13,060 queries (100%)	Overlapping Click Queries – 5072 queries (39%)			No Common Clicks 7988 (61%)
	Equal Click Queries – 3777 (29%)		Some Common Clicks 1295 (10%)	
	Single Identical Click 3737 (29%)	Multiple Identical Clicks 40 (< 1%)		
Equal Query Queries 4256 (33%)	Navigational Queries 3100 (24%)	36 (< 1%)	635 (5%)	485 (4%)
Different Query 8804 (67%)	637 (5%)	4 (< 1%)	660 (5%)	7503 (57%)

other behaviors. By adding restrictions, we reduce the number of false positives, and focus on instances with clear re-finding intent.

Repeat clicks are a reasonable proxy for re-finding intent. Thus, we are interested in the cases where users clicked on the same results during two different query instances:

1. *Overlapping-click queries* – Queries that have some common clicks. This type captures related intent and is the loosest form of repeated query. It is a superset of equal-click queries. Formally, given two click-through sets, C_1 and C_2 , corresponding to two queries, q_1 and q_2 , $C_1 \cap C_2 \neq \emptyset$.

Even assuming repeat clicks are a good representation for re-finding intent, overlapping-click queries do not necessarily reflect re-finding exclusively. Users may explore new results as well as old in overlapping-click queries (broadening their search). Or they may not want to re-find everything they found initially, but rather to concentrate on more specific sets of results (narrowing). The category of equal-click queries is more restrictive:

2. *Equal-click queries* – The user clicks on the same results for

the two queries. The queries may not be the same. Formally, for two queries, q_1 and q_2 , and the two corresponding click-through sets C_1 and C_2 , $C_1 = C_2$.

While looking at click patterns is likely to give a relatively accurate picture of whether or not the user is re-finding, search engines generally do not know what their users are going to click on at the time a query is issued. For this reason, we consider equivalence and similarity in the query strings themselves. In the general case we can ignore the clicks associated with each query.

3. *Equal-query queries* – The user issues the same query but visits a potentially disjoint set of Web pages. Given two queries, q_1 and q_2 , we have $q_1 = q_2$.

Clearly, a combination of the two dimensions represents a very narrow, but precise, definition of re-finding intent:

4. *Navigational queries* – Queries where the user makes the same query and always clicks to one and only one result are assumed to have a navigational intent. Given two queries, q_1 and q_2 , and two corresponding click-through sets, C_1 and C_2 , a navigational query is one in which $q_1 = q_2$, $C_1 = C_2$ and $|C_1| = |C_2| = 1$ (in practice, we find that when $C_1 = C_2$ the size of both is nearly always 1).

Navigational queries, as defined above, tended to be for specific corporate Websites, and were likely part of a daily routine or at least daily life. By far the largest category of navigational queries contained searches for stores or businesses. Seventy-five of the 3100 navigational queries (2.4%), contained the word “bank”, presumably issued by users accessing online banking. Two other similarly sized categories of navigational queries contained the word “news” (81/3100, 2.6%) and “mail” (80/3100, 2.6%).

An interesting category of re-finding queries was defined by the entry of a URL, or a portion of a URL, in the search box. Of the 617 unique navigational query instances, 69 (11%) included “.com” in the query string. These represent 550 of the 3100 total navigational queries, or 18%. Although in many cases these users could have entered the URL into the navigation box instead of the search box, this is a very clear example of re-finding behavior that needs to be supported by the search engine.

Currently browsers, search engine query boxes, and toolbars are designed to encourage navigational queries by supporting history-based auto-completion. This interface feature makes it very likely that a user attempting to re-find will issue duplicate query strings. Realistically, however, not all re-finding behavior is captured by repeat queries. For example, an additional 5% of queries (over the 24% of queries classified as navigational) contained different

Table 2: Ways that similar query strings can differ.

Difference	Example
<i>Exact</i>	“california secretary of state” and “california secretary of state”
<i>Capitalization</i>	“Air France” and “air france”
<i>Extra Whitespace</i>	“britney spears” and “britney spears”
<i>Word order</i>	“new york department of state” and “department of state new york”
<i>Stop words</i>	“atlas missouri” and “atlas of missouri”
<i>Non-alphanumerics</i>	“sub-urban” and “sub urban”
<i>Duplicate words</i>	“wild animal” and “wild wild animal”
<i>Word merge</i>	“wal mart” and “walmart”
<i>Domain</i>	“hotmail.com” and “hotmail”
<i>Stemming and Pluralization</i>	“island for sale” and “islands for sale”
<i>Words swaps</i>	“american embassay london” and “american consulate london”
<i>Add/Remove Word</i>	“orange county venues” and “orange county music venues”
<i>Add/Remove Noun Phrase or Location*</i>	“Wild Adventures in Valdosta Ga” and “Wild Adventures”
<i>Abbreviations*</i>	“ba” and “British Airways”
<i>Synonyms*</i>	“Practical Jokes” and “Pranks”
<i>Misspellings*</i>	“yahoo” and “yhao”
<i>Reformulations*</i>	“UN Secretary-General” and “Kofi Annan”

Table 3: The transformations that lead to the most frequent normalizations in different analyses.

Different Query-Single Identical Click (DQ-SIC) (311 instances)	Different Query-Overlapping Click (DQ-OC) (276)	Controlled Study (27 instances)
Word removal only (83 instances, 27%)	Word removal only (90, 8%)	Stemming only (4, 15%)
Capitalization only (51, 16%)	Word swap only (40, 6%)	Capitalization only (4, 15%)
Word swap only (25, 8%)	Capitalization only (22, 5%)	Capitalization and word swap (3, 11%)
Word merge only (22, 7%)	Non-alphanumeric removal and word removal (17, 3%)	Capitalization and word replacement (2, 4%)
Non-alphanumeric removal and stop word removal (14, 5%)	Non-alphanumeric removal and word swap (14, 3%)	14 other combinations at 1 instance each

query strings that produced a single equivalent click ($|C_1| = |C_2| = 1$ but $q_1 \neq q_2$). These queries are likely navigational in intent but do not fall under the navigational query category described above.

Below we consider normalization functions (functions that render two queries *normalization-equivalent*). These functions allow different queries with similar intent to be identified with each other. That is, even though $q_1 \neq q_2$, $n(q_1) = n(q_2)$.

4.1 How Queries Can Differ

Query strings used to re-find can differ from their original forms in many ways. It has been shown that traditional vector space measures of similarity are generally unsuitable for finding query similarity [15]. To understand how to identify re-finding, we explored a number of potential differences between similar queries, enumerated in Table 2. Most of the differences listed are trivial to identify automatically, but some are not. Those that are starred – including abbreviations, synonyms, and reformulations – are not considered in our analysis for this reason.

Note that normalization functions must be selected carefully because many queries that look similar represent searches without any overlapping clicks and thus are likely to be searches for which there was no re-finding intent. There is an obvious trade-off between the precision of the queries matched and the recall.

4.2 Normalization in the Logs

In order to find an appropriate query normalization function we began by concentrating on those queries that represent potential re-finding intent despite the query string being different. This class includes different query/single identical click (DQ-SIC) as well as the different query/overlapping click (DQ-OC) categories (see Table 1). The first category (DQ-SIC) is very likely to represent simple re-finding as the user travels directly to the same Webpage. The second category (DQ-OC) contains more possible variation in query intent, but is interesting to explore because search engines should support complex re-finding behavior in addition to the simple single repeat click behavior. Previous work has identified over-lapping click queries as likely to be related in meaning, and therefore useful for clustering queries [25].

To find the optimal normalization our system automatically tested all 2049 possible combinations of the 12 top transformations from Table 2 to find the minimal set of transformations that generated normalization-equivalence between each query pair. More than one transformation was often necessary to generate equivalence.

For DQ-SIC queries we were unable to find a simple mapping for 112 of the 423 unique instances (26%). These queries tended to be very different. For example, topically related terms with no words in common could result in the same click. Other queries required too many changes (e.g., abbreviations) and were considered failed inputs. Of the successful transformations, 111 (or 36% of 311) required the removal of one word (e.g. “disney world” and “walt disney world”). Similarly, 44 (14%) required a word to be swapped. Generally only one transformation was necessary to generate equivalent queries (79%).

For the unique DQ-OC queries, 142 of the 413 unique DQ-OC instances (34%) could not be normalized. Overlapping clicks are likely related to the distinctness of the query strings, as suggested by our difficulty normalizing DC-OC queries. The more distinct two queries are, the more likely they are to generate distinct result sets. We also find a higher incidence of word swapping (68/276 or 25%) and word removal (132/276 or 48%) in the case of DQ-OC than for DQ-SIC. Likewise, only 180 queries (65% of 276) could be normalized with only one step, a drop from the DQ-SIC case. Unsurprisingly, the lesson from this analysis is that queries which are exact or near repeats of previous queries are more likely to generate the exact same clicks as before.

Although such patterns observed in the logs imply re-finding intent, the intent is never explicit. To further explore query normalization we initiated a controlled study of volunteers doing an explicit re-finding task. The results, described below, allow us to further sharpen our understanding of re-finding and serve as a useful comparison to the log study.

4.3 Controlled Study: How People Remember

To better understand how people remember past queries, we analyzed the data collected through a separate university-based small-scale study where volunteer participants were asked to issue a self-selected query and interact with the search results as they normally would. After an interval of a half hour to an hour, participants were emailed a brief survey that asked them to remember the query they issued without referring back to it. The results of this study give insight into how easy it is to remember past queries and how likely people are to remember them.

One hundred and nineteen people participated in the study. Of those, 52% demographically self-identified as male, and 45% as female. Sixty four percent reported being aged 25 to 39, 18% over 40, and 15% under 25. Almost all (97%) reported using a computer daily. In general the follow-up survey was completed within a couple of hours of the initial search. Sixty-nine percent of all survey responses were received within three hours of the initial search, and all but five were received within a day. The average initial query length was 3.2 words, again, comparable to previous work [10]. Even though the elapsed time between a participant’s initial query and the moment when he or she was asked to remember it was relatively short, the original query was misremembered 30% of the time (36 of 119 query pairs).

We applied the same combinatorial analysis to the data collected through this study as we did to the query logs. Of the 36 misremembered queries, 27 (or 75%), were found to be equivalent after some normalizations. The nine remaining query pairs appear to have arisen from participants summarizing their previous queries instead of repeating them (e.g., “whats the best pricing available for a Honda Pilot or Accura MDX ?” → “best

pricing for Accura MDX”). Of the 27 that were normalization-equivalent, case normalization was needed in 12 cases (or 44%). In 9 of the 27 (33%), stemming of individual terms was necessary. In a full 30% (8/27) one word was substituted. For some this was due to misspellings (e.g., “helment” instead of “helmet”), and for others it was due to the use of synonyms (e.g., “where might I find...” versus “where can I find...”).

Interestingly, only four query pairs of the 27 (14%) were normalized by removal of a word. This is in contrast to the 36% in the log study (DQ-SIC case). Sixteen queries (or 59% of the 27) required more than one normalization step (e.g., word order *and* stop word removal).

The six most effective normalizations for our experiments are shown in Table 3. One notable difference between the two data sets is that the log contained many more instances of duplicate words and word ordering changes. There are a number of reasons why the anonymous log and controlled volunteer studies may differ. In particular, there was a significant difference in elapsed time between queries for the two studies. The average time between overlapping-click queries in the logs is over 12.2 days (292 hours) with a median of 30 hours. This longer interval presents many more opportunities for users to forget or change their queries. However, in the log study users often had the opportunity to learn effective querying for frequently sought information. Many log queries were issued more than twice and likely to be more memorable as a result. The controlled study was biased towards queries being memorable in that participants knew they were participating in a study, and against it in that the recall event was prompted by an email rather than self-directed.

5. INDIVIDUAL BEHAVIORS

Because navigational queries were prevalent, we explored their frequency and significance for individual users. Of the 114 users, 102 issued at least one equal-query query, and 87 (76%) performed at least one navigational query. Fifteen had equal-query queries but no navigational queries, possibly indicating they were explorers [26]. If we remove the bottom 25th percentile of users who issued the fewest queries (those issuing under 19 queries) we find that 76 (88%) of 86 have issued at least one navigational query (the distribution of these is shown in Figure 2).

Of the 87 users performing navigational queries, the median user issued 3 unique navigational queries. The average user issued 7.6 unique navigational queries (although removing one user, with a remarkable 103 unique navigational queries, decreases the average to 6.5 unique navigational queries). Considering those users who have issued navigational queries we find that on average ~10% of those users’ unique queries are navigational (median of 6%). This distribution is depicted in Figure 3.

Analysis of individual behavior may lend itself to the detection of robots and search engine optimizers. Users with many regularly spaced navigational queries are possibly using an automated system. For example, one trace contained 50 navigational queries

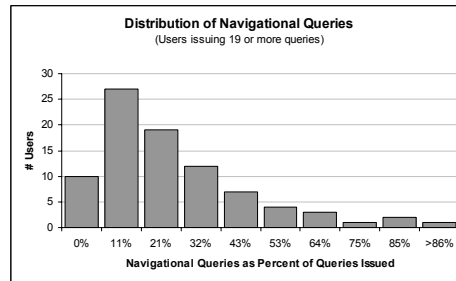


Figure 2: Distribution of Navigational Queries

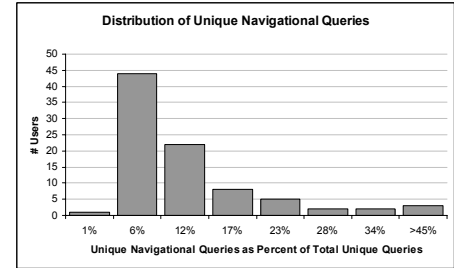


Figure 3: Distribution of unique navigational queries (for those users issuing navigational queries).

in 52 visits (96.2%); another contained 334 navigational queries out of 417 total queries (80.1%).

While it is clear from this analysis that not all users issue many re-finding queries, nearly all issue some and for many this is a significant portion of their query behavior.

6. AFFECT OF CHANGE ON RE-FINDING

Given the prevalence of re-finding queries, it is important to understand which search engine features help and which negatively impact the user’s re-finding objective. Search engines are constantly attempting to improve results through the discovery of new resources and the creation of new ranking strategies. While this benefits users who are looking for the best new information, the rank change of previously viewed search results can adversely impact those users attempting to re-find. Since the queries in our logs occurred over a sufficient period of time for the results to change in response to repeated queries, it was possible to observe to what extent changes to search result ranking affected the users’ ability to re-find information.

We found that when a previously clicked result changed position, users were less likely to re-click results. This suggests that changes to result ordering caused people to re-find less information and view more new information. This is not necessarily a bad thing if the new information is better than the old. However, users frequently would like to find the same result, as evidenced by the significant number of navigational queries. More critically, we observed that when the searcher clicked on a previously viewed result, the time it took to make the click was significantly longer if the rank of the result had changed in the meantime. This suggests that changing the rank of a result can lead to noticeable changes in user behavior; whether or not such changes are beneficial to the user should be considered carefully.

6.1 Rank Change Reduces Chance of Click

To understand how a change in a result’s rank affected click behavior, we looked at how likely a result was to be clicked again. Because the dataset did not contain results that were not clicked, we were only able to identify result lists that had changed when we observed rank changes among clicked results for queries with overlapping-clicks. A better understanding could be derived from a knowledge of which results were displayed, even if not clicked.

We looked at all queries that had overlapping-clicks. We compared the probability that any given click would be a repeat click for these queries under two conditions: (i) when a change in rank was observed among one of the overlapping clicks and (ii) where no rank change was observed. We found that it was much

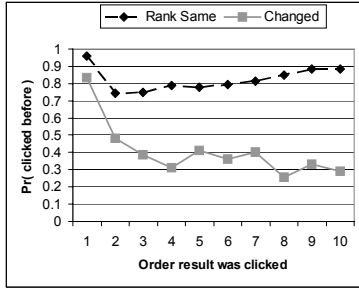


Figure 4: The probability of a result being clicked again as a function of the order in which the result was clicked.

more likely for a repeat result to be clicked if there was no change in rank: 88% percent of the clicks for overlapping-click queries were repeat clicks if there was no change in rank, while only 53% of the clicks were repeat clicks if there was a change in rank. These estimates were obtained by averaging all consecutive pairs of overlapping-click queries.

Figure 4 shows the probability that a clicked result was a repeat click as a function of the order in which the click occurs following a repeat overlapping-click query. The dashed curve corresponds to the probability averaged over those searches where no rank change was observed; the solid curve corresponds to an average where at least one result changed rank. Comparing the two curves, we see that a change in rank between queries makes it substantially less likely that a given result will be clicked on again during a follow-up search.

Also in Figure 4 we see a sharp drop in the probability of a repeat click between the first result and the second. Given that a finite number of results were clicked initially, it seems reasonable that if the user first clicks on repeat results then the probability of a repeat click would tend to drop with increasing numbers of clicks, as the user exhausts the set of previously-clicked results. The drop continues past click two when restricted to clicks on results with rank changes, which would seem to indicate that users are more likely to click on new results as they continue to interact with the result list than they are to click on previously clicked results which have changed rank.

It is not immediately obvious from this analysis whether a decreased likelihood of re-finding reflects a positive or negative influence of result list changes on user experience. It could be that the changes interfered with re-finding, or it could be that the searcher found new and better information in the new result set.

6.2 Rank Change Slows Re-Finding

To get a better idea of whether changes interfered with re-finding, we looked at queries where we were certain that information was being re-found, as evidenced by a repeat click. Because easy searches are likely to take less time than harder searches, we looked at the time interval between a search and a click on a result that was seen before. For this reason, we measured the time from when a query was issued until the common URL was clicked for different-query, overlapping click queries.

Table 4 shows the average number of seconds it took to click a URL that was clicked during the initial session when that URL was (i) shown at the same rank it originally appeared, and (ii) shown at a different rank. If the rank of the result had not

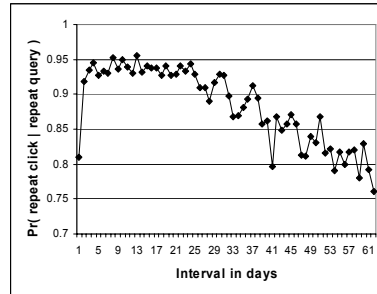


Figure 5: Probability of repeat click

changed, the second click occurred relatively quickly, while if the rank had changed, it took significantly ($p < 0.01$) longer. Changes to result ordering appear to slow re-finding.

A natural question is whether a positive rank change (a result moving up) or a negative rank change (a result moving down) impact search time differently. Because our log does not contain a significant number of rank changes of each type, it is difficult to make an argument of statistical significance. However, the data is suggestive of a positive improvement in time-to-click for positive changes in rank as well as some benefit to no change (likely due to learning). When previously clicked results move down in rank, time-to-click increases. A hypothesis consistent with previous work on eye-tracking

in search [7, 8] is that users pay more attention to early-ranked items. Thus, if a previously clicked on result moves up, it is more easily re-found via a visual scan. In the future, we hope to statistically confirm these findings by using longer traces.

7. PREDICTING THE QUERY TARGET

A potentially desirable search engine behavior, given the impact of rank changes, is to impose more stability on the results returned by searches where re-finding is deemed to be the intent. To do this, it is necessary to quickly and accurately classify queries to determine the best results for a given user given re-finding intent. This section looks at predicting whether a previously viewed result will be clicked based on the query string and past clicks.

Repeat searchers may be looking for new information, or they may want information that they have seen before. It was most common to look for the same information: approximately 87% (3692/4256) of equal-query queries were also overlapping-click queries. Fewer queries (1632 or 38%) resulted in at least one unique click. Searchers did not always want only old or only new information when they issued equal-query queries, as 25% of the queries, or 1070, involved both a repeat click and a unique click.

This section begins by looking at the effect that elapsed time and number of previous clicks have on repeat queries. Navigational queries are particularly easy to predict, and they are discussed in greater detail, as are other query types.

7.1 The Effect of Elapsed Time

We looked at how the elapsed time between equal-query queries affected the likelihood of observing a repeat click. The probability of a repeat click as a function of elapsed time between identical queries can be seen in Figure 5. Repeat queries that were issued very close together in time (e.g., within several hours) had a relatively low probability of resulting in a repeat click. The probability of a repeat click for queries re-issued within an hour is

Table 4: Time-to-click (in seconds) as a function of rank change.

Query type	Mean	Median	StdDev
All (rank changed & unchanged)	178	22	333
• Rank unchanged (all types)	94	6	234
• Rank changed (all types)	192	26	365
• <i>Equal-query</i>	186	22	343
• <i>Non-equal</i>	147	25	288
• <i>Non-equal-no rank change</i>	43	6	94
• <i>Non-equal -rank change</i>	226	77	354

64%, compared with the earlier reported overall average of 87%. Queries repeated very quickly probably occurred as part of the same search session, and represent instances where the user was looking for something new.

The probability of repeat clicks climbs quickly, however, for intervals longer than a day or two. Once it reaches a peak, the probability of a repeat click between identical queries slowly declines. This may represent a trend to forget previously seen information over time.

7.2 Navigational Queries

We found we were able to accurately predict the likelihood of a repeat click based on a history of clicked results from past queries. Navigational queries were particularly easy to predict. Recall that navigational queries are equal-query queries where the user clicked on the same result for each query instance and did not click on any other results. Using this definition, 507 (or 47%) of all unique equal-query queries issued were labeled navigational.

Navigational queries tended to be somewhat shorter in length than other queries (13.6 characters, compared with 16.4 characters for non-navigational equal-query queries and 16.7 characters for overlapping-click queries). This seems reasonable because navigational queries are probably intended to be an easy way to return to a Web page, and thus should be short and easy to remember. Navigational queries were also more likely to include an indication that they were a search for a URL: 12% of all navigational queries contained “.com”, “.edu”, or “.net”, as opposed to only 5% of non-navigational equal-query queries.

Navigational queries were also repeated more often than other repeat queries (4.0 times, compared with 3.8 for equal-query queries and 3.3 for overlapping-click queries) and, as found by Sanders and Dumais [18], the interval between navigational queries was longer (22 days, compared with 20 days and 16 for equal-query and overlapping-click queries respectively). It is likely that navigational queries occurred more often because they are more of an access than a search strategy, and people tend to access more than search. The longer intervals are probably because the queries are probably chosen to be particularly memorable even across long periods of time.

It was easy to predict whether a query was navigational given two previous instances of the same query as training data. By doing this, we were able to automatically classify 1841, or 12%, of all observed searches on the fly as navigational. For these searches, we could predict with 96% accuracy one of the URLs clicked. When restricted to predicting the first URL clicked, accuracy only dropped slightly, to 95%, and if predicting that only that URL was clicked, accuracy dropped slightly more, to 94%.

It was less easy to identify a navigational query using only one previous query. While doing so covers more of the data (2955, or 23%, of the searches), the prediction was right only 87% of the time. Given 87% of all equal-query queries involve overlapping clicks, it is not at all surprising that we can predict exactly which result will be clicked 87% of the time given we know the user only clicked one result before.

7.3 Other Types of Repeat Queries

We also investigated whether it was possible to predict whether a user was going to click on new or repeat results for equal-query queries that were not navigational. Using features suggested by

the earlier analysis presented in this paper, such as elapsed time, query length, and number of results clicked previously, we trained an SVM (http://www.cs.cornell.edu/People/tj/svm_light/) to predict two outcomes: (i) whether or not a new result would be clicked, and (ii) whether or not a repeat result would be clicked.

The strongest predictors for a click on a new result included the number of times the query was issued previously (and if it was issued more than once before), whether any previously viewed result was clicked more than once, and several features that were the same for queries that were repeated only twice:

- Number of clicks the first time the query was issued
- Number of clicks the previous time the query was issued
- Number of unique clicks the previous time

While no correlation was found between the number of clicks and the likelihood of a repeat click, given the value of these features in predicting new clicks it seems it is indicative of a new click.

The strongest predictors for a repeat click were a) that only one result was clicked during the previous search, and b) that the query had been issued more than once. These features are also useful for identifying navigational queries, which experience a high incidence of repeat queries (although queries identified as navigational queries were excluded from this analysis).

Using the features described above, and leave-one-out cross-validation, we compared the ability of the SVM to predict whether a new result or a unique result would be clicked. As our baseline we used the accuracy that could be expected if people were always assumed not to click on something new (61.4% accuracy) and to click on something they clicked before (74.7% accuracy). In both cases, we found the SVM was able to make a significantly ($p < 0.01$) better prediction at 79.3% accuracy for new clicks (an increase of 30%), and 78.1% for repeat clicks (an increase of 5%). The SVM probably does a better job predicting new clicks than old because the navigational query data, which was the most easily identifiable repeat click data, was excluded.

We also looked at including the user as a feature in the learning. While this led to a slight improvement in both cases (80.1% accuracy in predicting new clicks and 79.4% accuracy in predicting repeat clicks), the difference was not significant. However, we suspect that users do exhibit distinct repeat and new click behaviors, and we probably need to accumulate additional features that will allow us to capture this distinction.

8. DESIGN IMPLICATIONS

The findings presented in this paper have many ramifications for search engine design, and potentially for browsers and search toolbars. Re-finding, or searching for previously found information, represents a significant fraction of user behavior. Traditionally, search engines have focused on returning search results without consideration of the user's past query history, but the results of the log study suggest it might be a good idea for them to do otherwise. Although finding and re-finding tasks may require different strategies, tools will need to seamlessly support both activities. As shown in the log analysis, people often clicked on both old and new results during the same search.

Because people repeat queries so frequently, search engines should assist their users by providing a means of keeping a record of individual users' search histories, perhaps via software installed on the user's own machine. A number of search history designs

have been explored (e.g., [12]). The results of the log study indicate it is important to account for individual differences in how people repeat queries. For example, different users made use of repeat queries at different rates, and may benefit from having a different amount of screen real estate devoted to displaying their search history. Furthermore, search histories could be customized based on many factors including the time of day. Users with a large number of navigational queries may also benefit from the direct linking to the Webpage (possibly labeled with the frequent query term). This form of shortcut could be highly effective for many in terms of rapid access to information.

While a user may simultaneously have a finding and re-finding intent when searching, satisfying both needs may be in conflict. Finding new information means being returned the best new information, while re-finding means being returned the previously viewed information. We found that when previously viewed search results changed to include new information, the searcher's ability to re-find was hampered. It is important to consider how the two search modalities can be reconciled so a user can interact with new, and previously seen, information. As Teevan [20] has previously proposed, before information is allowed to change, it is important to understand which aspects of it that a person has already interacted with are memorable.

Despite the personal nature of re-finding, it is possible that repeat queries from one user could benefit another. For example, popular results for navigational queries could be globally elevated by the search engine for the benefit of everyone. While desirable in theory, in practice this may encourage search engine spam. In contrast, personalizing search results based on search history can help avoid potential problems caused by spam.

9. CONCLUSION

In this paper, we looked at the queries issued to a leading Internet search engine in 114 user search traces over the course of a year, and studied 119 users in a separate university-based controlled experiment with volunteers. We observed that repeat searches and repeat clicks were very common. We found it was possible to predict which queries were navigational and what results were likely to be clicked. Changing the rank of a previously clicked result appears to hinder re-finding, so click predictions should be used carefully by search engines to customize search results in a manner consistent with the user's search habits. We are currently continuing work in this area with a larger set of users. In particular we are interested in further analyzing repeat queries for individual users and broader notions of repetition (e.g., repeating query chains, or co-occurrences of queries in time). We are also pursuing an understanding of user behavior during the potentially iterative process of refining a query for re-finding tasks.

10. REFERENCES

- [1] Aula, A., Jhaveri, N., and Käki, M. Information search and re-access strategies of experienced Web users. In *Proceedings of WWW '05*, 2005, 583-592.
- [2] Beitzel, S. M., Jensen, E. C., Chowdhury, A., Grossman, D. and Frieder, O. Hourly analysis of a very large topically categorized Web query log. In *Proceedings of SIGIR '04*, 2004, 321-328.
- [3] 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.
- [4] Capra, R. and Pérez-Quñones, M.A. Using Web search engines to find and re-find information. *IEEE Computer*, 38 (10), 2005, 36-42.
- [5] Cockburn, A., Greenberg, S., Jones, S., Mckenzie, B. and Moyle, M. Improving Web page revisitation: Analysis, design and evaluation. *IT & Society*, 1 (3), 2003, 159-183.
- [6] Dumais, S. T., Cutrell, E., Cadiz, J. J., Jancke, G., Sarin, R. and Robbins, D. C. Stuff I've Seen: A system for personal information retrieval and re-use. In *Proceedings of SIGIR '03*, 2003, 72-79.
- [7] Enquiro. Did-it, Enquiro, and Eyetools uncover search's golden triangle. www.enquiro.com/eye-tracking-pr.asp (last retrieved 9/29/2006)
- [8] Granka, L.A., T. Joachims, and Gay, G. Eye-Tracking Analysis of User Behavior in WWW Search," In *Proceedings of SIGIR '04*, 2004, 478-479.
- [9] Graphic, Visualization, and Usability Center. Gvu's Tenth WWW User Survey, October 1998.
- [10] Jansen, B. J., Spink, A. and Saracevic, T. Real life, real users, and real needs: A study and analysis of user queries on the Web. *Information Processing and Management*, 36 (2), 2000, 207-227.
- [11] Jones, R. and Fain, D. C. Query word deletion prediction. In *Proceedings of SIGIR '03*, 2003, 435-436.
- [12] Komlodi, A., Soergel, D., and Marhionini, G. Search histories for user support in user interfaces. *JASIST*, 57(6): 803-807, 2006.
- [13] Lau, T. and Horvitz, E. Patterns of search: Analyzing and modeling Web query refinement. In *Proceedings of the UM '99*, 1999, 119-128.
- [14] Obendorf, H., Weinreich, H., Herder, E., and Mayer, M. Web page revisitation revisited: Implications of a long-term click-stream study of browser usage. In *Proceedings of CHI '07*, 2007, 597-606.
- [15] Raghavan, V. and Server, H. On the reuse of past optimal queries. In *Proceedings of SIGIR '95*, 1995, 344-350.
- [16] Rainie, L., and Shermak, J. *Search engine use November 2005*, Pew Internet & American Life Project, Washington DC, 2005.
- [17] Ross, N. C. M. and Wolfram, D. End user searching on the Internet: An analysis of term pair topics submitted to the Excite search engine. *JASIST*, 51 (10), 2000, 949-958.
- [18] Sanderson, M. and Dumais, S. Examining repetition in user search behavior. In *Proceedings of ECIR '07*, 2007.
- [19] Tauscher, L. and Greenberg, S. How people revisit Web pages: Empirical findings and implications for the design of history systems. *International Journal of Human-Computer Studies*, 47 (1), 1997, 97-137.
- [20] Teevan, J. Supporting finding and re-finding through personalization. Doctoral Thesis, MIT, February 2007.
- [21] Teevan, J., Adar, E., Jones, R. and Potts, M. History repeats itself: Repeat queries in Yahoo's logs. In *Proceedings of SIGIR '06*, 2006, 703-704.
- [22] Teevan J., Alvarado C., Ackerman M. S., and Karger D. R. The perfect search engine is not enough: A study of orienteering behavior in directed search. In *Proceedings of CHI '04*, 2004, 415-422.
- [23] Wang, P., Berry, M. W. and Yang, Y. Mining longitudinal Web queries: Trends and patterns. *JASIST*, 54 (8), 2003, 743-758.
- [24] Wedig S. and Madani, O. A large-scale analysis of query logs for assessing personalization opportunities. In *Proceedings of KDD '06*, 2006, 742-747.
- [25] Wen, J.-R., Nie, J.-Y. and Zhang, H.-J. Query clustering using user logs. *TOIS*, 20 (1), 2002, 59-81.
- [26] White, R. and Drucker, S. M. Investigating behavioral variability in Web search. In *Proceedings of WWW '07*, 2007, 21-30.